



\* 7332 10056: contig of 2725 bp in length  
\* 10057 10156: gap of unknown length  
\* 10157 13400: contig of 3244 bp in length  
\* 13401 13500: gap of unknown length  
\* 13501 16437: contig of 2937 bp in length  
\* 16438 16538: gap of unknown length  
\* 20357 20456: contig of 3819 bp in length  
\* 20357 20456: gap of unknown length  
\* 20357 23304: contig of 2848 bp in length  
\* 23305 23404: gap of unknown length  
\* 23405 26016: contig of 2612 bp in length  
\* 26017 26116: gap of unknown length  
\* 26117 31388: contig of 5272 bp in length  
\* 31389 31489: gap of unknown length  
\* 31489 37603: contig of 6115 bp in length  
\* 37604 37703: gap of unknown length  
\* 37704 45366: contig of 7663 bp in length  
\* 45367 45466: gap of unknown length  
\* 45467 51573: contig of 6107 bp in length  
\* 51574 51674: gap of unknown length  
\* 51674 58444: contig of 6771 bp in length  
\* 58445 58544: gap of unknown length  
\* 58545 67500: contig of 8956 bp in length  
\* 67501 67600: gap of unknown length  
\* 67601 78970: contig of 11370 bp in length  
\* 78971 79070: gap of unknown length  
\* 79071 86855: contig of 7785 bp in length  
\* 86856 86956: gap of unknown length  
\* 86957 98790: contig of 11835 bp in length  
\* 98791 98890: gap of unknown length  
\* 98891 125623: contig of 26733 bp in length  
\* 125624 125723: gap of unknown length  
\* 125724 157663: contig of 31939 bp in length  
\* 157663 182356: gap of unknown length  
\* 182356 182455: contig of 24593 bp in length  
\* 182456 220367: contig of 37912 bp in length.

FEATURES  
source  
1. 220367  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-133E17"  
/clone\_lib="RPCI mouse BAC library 23"

ORIGIN  
Query Match 71.4%; Score 20; DB 2; Length 220367;  
Best Local Similarity 82.1%; Pred. No. 75;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACTC 28  
Db 194657 GCTGAGGCTAAGCTACACAGTACGACTC 194630

Fragment Name Begin End  
LMFLCHR31\_00 1 110000  
LMFLCHR31\_01 100001 210000  
LMFLCHR31\_02 200001 310000  
LMFLCHR31\_03 300001 410000  
LMFLCHR31\_04 400001 510000  
LMFLCHR31\_05 500001 610000  
LMFLCHR31\_06 600001 710000  
LMFLCHR31\_07 700001 810000  
LMFLCHR31\_08 800001 910000  
LMFLCHR31\_09 900001 1010000  
LMFLCHR31\_10 1000001 1110000  
LMFLCHR31\_11 1100001 1210000  
LMFLCHR31\_12 1200001 1310000

Sequence split into 22 fragments LOCUS LMFLCHR31 Accession AL499621

RESULT 8  
WPCOMMENT-  
LMFLCHR31\_01

LMFLCHR31\_13 1300001 1410000  
LMFLCHR31\_14 1400001 1510000  
LMFLCHR31\_15 1500001 1610000  
LMFLCHR31\_16 1600001 1710000  
LMFLCHR31\_17 1700001 1810000  
LMFLCHR31\_18 1800001 1910000  
LMFLCHR31\_19 1900001 2010000  
LMFLCHR31\_20 2000001 2110000  
LMFLCHR31\_21 2100001 2117963

Continuation (2 of 22) of LMFLCHR31 from base 100001 (AL499621 Leishmania major chromosome)

Query Match 70.7%; Score 19.8; DB 2; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 95;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTAC 23  
Db 50914 GCAGTGGCGCAGCTCCACAGTAC 50936

RESULT 9  
HSA322632  
LOCUS  
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone  
ACCESSION AJ322632  
VERSION AJ322632.1 GI:15867011  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 890)  
AUTHORS Karsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M., Podovsky, R.M., Matshukin, Y.G., Gvanchandani, A., Muravenko, O.V., Levitsky, V.G., Kolchanov, A.I., Protopopov, A.I., Kashuba, V.I., Kisselev, L.L., Wasserman, W., Wahlstedt, C. and Zabarovsky, E.R.  
TITLE NotI flanking sequences: a tool for gene discovery and verification of the human genome  
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)  
MEDLINE 22131767  
PUBMED 12136098  
REFERENCE 2 (bases 1 to 890)  
AUTHORS Zabarovsky, E.R.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden

FEATURES  
source  
1. 890  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="NLI-HB3C"

ORIGIN  
Query Match 70.0%; Score 19.6; DB 9; Length 890;  
Best Local Similarity 84.6%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGGCGCAGCTACACAGTACGACTC 28  
Db 737 TGTGTGGAGCTACACAGGACGTCTC 762

RESULT 10  
BX248512  
LOCUS  
DEFINITION Danio rerio clone CH211-284D12, WORKING DRAFT SEQUENCE, 6 unordered pieces.  
ACCESSION BX248512  
VERSION BX248512.2 GI:30230825  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.



SOURCE  
ORGANISM Danio rerio (zebrafish)

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
JOURNAL Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 157060)

COMMENT  
McLaren S.  
Direct Submission  
Submitted (20-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 29, 2003 this sequence version replaced gi:28316003.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zC284D12  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 154848 bases at least Q40  
Consensus quality: 155730 bases at least Q30  
Consensus quality: 156280 bases at least Q20  
Insert size: 156560; sum-of-contigs  
Insert size: 154730; 5.7% error; agarose-fp  
Quality coverage: 6.62x in Q20 bases; sum-of-contigs Quality  
coverage: 8.25x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\*  
\* 1 27215: contig of 27215 bp in length  
\* 27316: gap of 100 bp  
\* 75554: contig of 48239 bp in length  
\* 75555: gap of 100 bp  
\* 76555: contig of 2751 bp in length  
\* 78406: gap of 100 bp  
\* 78506: contig of 2976 bp in length  
\* 81481: gap of 100 bp  
\* 81482: contig of 2509 bp in length  
\* 81582: gap of 100 bp  
\* 84091: contig of 72870 bp in length.  
\* 84191 157060: contig of 72870 bp in length.  
-----  
Location/Qualifiers  
1. 157060  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-284D12"  
/clone\_lib="CHORI-211"  
1. 27215  
/note="assembly fragment:01021"  
fragment chain:1  
clone end:SP6  
vector\_side:left  
27316..75554  
/note="assembly fragment:01284"  
fragment chain:1  
75555..78406  
/note="assembly fragment:00233"  
78506..81481  
/note="assembly fragment:00575"  
81582..84090  
/note="assembly fragment:00895"  
84191..157060  
/note="assembly fragment:01502"  
clone\_end:T7

ORIGIN  
vector\_side:right

Query Match 70.0%; Score 19.6; DB 2; Length 157060;  
Best Local Similarity 84.6%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTGTGGCGGCTACACAGTACGAC 26  
|||||  
Db 77166 GCTGTGGCGGCTACATATGAC 77191  
|||||

RESULT 11  
AC128968 158106 bp DNA linear HTG 15-NOV-2002  
LOCUS Ratcus norvegicus clone CH230-357E12, WORKING DRAFT SEQUENCE, 2  
DEFINITION unordered pieces.  
AC128968  
AC128968.3 GI:25007346  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Ratcus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 158106)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alabrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, B., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunsaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hernandez, J.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoggues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, H., Lou, L., Lou, L., Lozada, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Mallory, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,  
Nwankwelu, O., Okwuonu, G., Olarnpungsoo, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Platter, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. B.,  
Sanders, W., Savary, G., Scherer, S., Scott, G., Shateman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Walker, B., Wang, J.,  
Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wiczyski, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von





as. compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEY-27M7 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.  
Location/Qualifiers  
1. 234545  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-27M7"  
/clone\_lib="DanioKey"

## ORIGIN

Query Match 70.0%; Score 19.6; DB 5; Length 234545;  
Best Local Similarity 84.6%; Pred. No. 1.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTGTGGCGAGCTACAGATGAC 26  
|||||  
Db 111104 GCTGTGGCGAGCTACAGATGAC 111079

## RESULT 15

AC106176 239330 bp DNA linear HTG 13-MAY-2003  
LOCUS Rattus norvegicus clone CH230-20P16, WORKING DRAFT SEQUENCE, 3  
unordered pieces.  
AC106176  
AC106176.5 GI:30578781  
HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

1 (bases 1 to 239330)  
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Dwy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gili, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelemeh, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## Direct Submission

Unpublished  
2 (bases 1 to 238330)

Worley, K.C.

## Direct Submission

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 238330)

Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:24819215.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## Genome Center

Center: Baylor College of Medicine

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Center project name: GJOH

Center Clone name: CH230-20P16

Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 229461 bases at least Q40

Consensus quality: 231714 bases at least Q30

Consensus quality: 233148 bases at least Q20

Estimated insert size: 239075; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\*

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 173.353 Seconds

(without alignments)  
514.627 Million cell updates/sec

Title: US-10-624-714-6

Perfect score: 21

Sequence: 1 gattgcgcagcgcacacacgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

1: Geneseqn1990s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2000s:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002s:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21	100.0	110000	4	AAI99682_04
C 2	21	100.0	110000	4	AAI99683_04
C 3	18.4	87.6	412	5	AAF66212
C 4	16.8	80.0	599	7	ABZ55428
C 5	16.8	80.0	867	7	ACF70867
C 6	16.8	80.0	2423	3	ACN98795
C 7	16.8	80.0	2894	4	ABL21857
C 8	16.8	80.0	5850	4	ABL21632
C 9	16.8	80.0	6304	4	ABL12856
C 10	16.8	80.0	110000	7	ACF67367_39
C 11	16.8	80.0	110000	7	ACF65388_08
C 12	16.2	77.1	286	6	ABN22495
C 13	16.2	77.1	505	7	ABZ52448
C 14	16.2	77.1	1488	7	ACA25079
C 15	16.2	77.1	2847	4	AAI36722
C 16	16.2	77.1	2847	4	AAI36721
C 17	16.2	77.1	2847	7	ABX59709
C 18	16.2	77.1	2847	7	ABX59710
C 19	15.8	75.2	763	6	ABX76772
C 20	15.8	75.2	3018	4	ABL25988
C 21	15.8	75.2	12483	7	ACF68520
C 22	15.8	75.2	110000	7	ACF67367_12
C 23	15.8	75.2	182624	7	ACF65379

## ALIGNMENTS

RESULT 1  
AAI99682\_04/c

Continuation (5 of 45) of AAI99682 from base 400001 (Mycobacterium tuberculosis strain H: WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

Fragment Name	Begin	End
AAI99682_00	100001	110000
AAI99682_01	200001	210000
AAI99682_02	300001	310000
AAI99682_03	400001	410000
AAI99682_04	500001	510000
AAI99682_05	600001	610000
AAI99682_06	700001	710000
AAI99682_07	800001	810000
AAI99682_08	900001	910000
AAI99682_09	1000001	1010000
AAI99682_10	1100001	1110000
AAI99682_11	1200001	1210000
AAI99682_12	1300001	1310000
AAI99682_13	1400001	1410000
AAI99682_14	1500001	1510000
AAI99682_15	1600001	1610000
AAI99682_16	1700001	1710000
AAI99682_17	1800001	1810000
AAI99682_18	1900001	1910000
AAI99682_19	2000001	2010000
AAI99682_20	2100001	2110000
AAI99682_21	2200001	2210000
AAI99682_22	2300001	2310000
AAI99682_23	2400001	2410000
AAI99682_24	2500001	2510000
AAI99682_25	2600001	2610000
AAI99682_26	2700001	2710000
AAI99682_27	2800001	2810000
AAI99682_28	2900001	2910000
AAI99682_29	3000001	3010000
AAI99682_30	3100001	3110000
AAI99682_31	3200001	3210000
AAI99682_32	3300001	3310000
AAI99682_33	3400001	3410000
AAI99682_34	3500001	3510000
AAI99682_35	3600001	3610000
AAI99682_36	3700001	3710000
AAI99682_37	3800001	3810000
AAI99682_38	3900001	3910000
AAI99682_39	4000001	4010000
AAI99682_40	4100001	4110000

ACD44940 Human SR-  
ABAI14114 Human ner  
ABAI1404 Human ner  
AAS52283 E. coli D  
ACA33247 Prokaryot  
ACA5573 Nucleotid  
AAB64856 Bordetell  
ABZ40086 N. gonorr  
ABZ41323 N. gonorr  
AAS1957 DNA encod  
AAX24498 Human SR-  
AAX24562 Human SR-  
AAX24554 Human SR-  
AAX24590 Human SR-  
ACD45017 Human SR-  
ACD44927 Human SR-  
ABA21206 Human ner  
AAH67049 C Glutami  
AAF72045 Corynebact  
ACA51845 Prokaryot  
ACA51106 Prokaryot  
AAS80001 DNA encod

WP AAI99682\_41 4100001 4210000  
WP AAI99682\_42 4200001 4310000  
WP AAI99682\_43 4300001 4410000  
WP AAI99682\_44 4400001 4411529

Query Match 100.0%; Score 21; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21

Db 72802 GATTGGCAGCGCCCAACAGC 72782

## RESULT 2

Continuation (5 of 44) of AAI99683 from base 400001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

Fragment Name	Begin	End
WP AAI99683_00	1	110000
WP AAI99683_01	100001	210000
WP AAI99683_02	200001	310000
WP AAI99683_03	300001	410000
WP AAI99683_04	400001	510000
WP AAI99683_05	500001	610000
WP AAI99683_06	600001	710000
WP AAI99683_07	700001	810000
WP AAI99683_08	800001	910000
WP AAI99683_09	900001	1010000
WP AAI99683_10	1000001	1110000
WP AAI99683_11	1100001	1210000
WP AAI99683_12	1200001	1310000
WP AAI99683_13	1300001	1410000
WP AAI99683_14	1400001	1510000
WP AAI99683_15	1500001	1610000
WP AAI99683_16	1600001	1710000
WP AAI99683_17	1700001	1810000
WP AAI99683_18	1800001	1910000
WP AAI99683_19	1900001	2010000
WP AAI99683_20	2000001	2110000
WP AAI99683_21	2100001	2210000
WP AAI99683_22	2200001	2310000
WP AAI99683_23	2300001	2410000
WP AAI99683_24	2400001	2510000
WP AAI99683_25	2500001	2610000
WP AAI99683_26	2600001	2710000
WP AAI99683_27	2700001	2810000
WP AAI99683_28	2800001	2910000
WP AAI99683_29	2900001	3010000
WP AAI99683_30	3000001	3110000
WP AAI99683_31	3100001	3210000
WP AAI99683_32	3200001	3310000
WP AAI99683_33	3300001	3410000
WP AAI99683_34	3400001	3510000
WP AAI99683_35	3500001	3610000
WP AAI99683_36	3600001	3710000
WP AAI99683_37	3700001	3810000
WP AAI99683_38	3800001	3910000
WP AAI99683_39	3900001	4010000
WP AAI99683_40	4000001	4110000
WP AAI99683_41	4100001	4210000
WP AAI99683_42	4200001	4310000
WP AAI99683_43	4300001	4403765

Query Match 100.0%; Score 21; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21

Db 72806 GATTGGCAGCGCCCAACAGC 72866

RESULT 3  
AAAF6212/c  
ID AAF6212 standard; cDNA; 412 BP.

XX AC AAF6212;

XX DT 09-APR-2001 (first entry)

XX DE Novel human polynucleotide, SEQ ID NO: 1968.

XX KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
XX KW breast cancer; lung cancer; cancer detection; ss.

XX OS Homo sapiens.

XX FN WC200102568-A2.

XX PD 11-JAN-2001.

XX PF 30-JUN-2000; 2000WO-US018374.

XX PR 02-JUL-1999; 99US-0142310P.

XX PR 02-JUL-1999; 99US-0142311P.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;  
XX PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;  
XX PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;  
XX PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;  
XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a  
XX mammalian cell and detecting cancer, particularly of the colon or  
XX prostate, comprises 3351 human polynucleotide sequences.

XX Claim 9; Page 827-828; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human  
XX polynucleotides. The library is used to detect differentially expressed  
XX genes correlated with a cancerous state of a mammalian cell and can  
XX detect colon, prostate, breast and lung cancer. The library can be used  
XX to produce probes for detection of mRNA and to produce additional copies  
XX of the polynucleotides. The probes can be used for chromosome mapping of  
XX the polynucleotide and for detection of transcription levels. Ribozymes  
XX or antisense oligonucleotides can be generated. The polynucleotides and  
XX their gene products are used as genetic or biochemical markers (e.g. in  
XX blood or tissues) that will detect the earliest changes along the  
XX carcinogenesis pathway and/or monitor the efficacy of therapies and  
XX preventive interventions. The polynucleotides, polypeptides and  
XX antibodies against them can be used in pharmaceutical compositions to  
XX treat the cancers and proliferative disorders such as neoplasia,  
XX dysplasia and hyperplasia

XX SQ Sequence 412 BP; 128 A; 104 C; 78 G; 101 T; 0 U; 1 Other;

Query Match 87.6%; Score 18.4; DB 5; Length 412;  
Best Local Similarity 95.0%; Pred. No. 35;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAG 20

Db 191 GATTGGCAGCGCCCAACAG 172

## RESULT 4

ABZ55428

ID ABZ55428 standard; cDNA; 599 BP.

XX AC ABZ55428;

XX





PN WO20005320-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US005989.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 PI WPI; 2000-579444/54.  
 XX  
 DR P-PSDB; AAB54030.  
 DR  
 DR  
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition.  
 XX  
 XX Claim 1; Page 511-512; 1379pp; English.  
 XX  
 XX AAC998773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, nontoxic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiac and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 2423 BP; 613 A; 616 C; 518 G; 658 T; 0 U; 18 Other;  
 Query Match 80.0%; Score 16.8; DB 3; Length 2423;  
 Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GATTGGCGAGCGCCACACAG 20  
 |||||  
 Db 1005 GATTGGCGAGCACCCTACAG 986  
 RESULT 7  
 ABL12857  
 ID ABL12857 standard; cDNA; 2894 BP.  
 XX  
 AC ABL12857;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33053.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.  
 XX  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 DR P-PSDB; ABB68754.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 PT  
 PS Claim 1; SEQ ID NO 33053; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2894 BP; 757 A; 763 C; 713 G; 661 T; 0 U; 0 Other;  
 Query Match 80.0%; Score 16.8; DB 4; Length 2894;  
 Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GATTGGCGAGCGCCACACAG 20  
 |||||  
 Db 608 GATTGGCGAGCGCTCACAG 627  
 RESULT 8  
 ABL21632  
 ID ABL21632 standard; DNA; 5850 BP.  
 XX  
 AC ABL21632;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16369.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 DR New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PT interactions.

XX Claim 1; SEQ ID NO 16369; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 5850 BP; 1632 A; 1568 C; 1264 G; 1386 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 4; Length 5850;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGCGCAGCGCCCAACAG 20

Db 5800 GATTTCGAGCGCCCAACAG 5819

RESULT 9

ABL12856/c

ID ABL12856 standard; cDNA; 6304 BP.

XX AC ABL12856;

XX 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 33050.KW *Drosophila*; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX *Drosophila melanogaster*.

XX WO200171042-A2.

PD 27-SEP-2001.

FF 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB68753.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions.

XX Claim 1; SEQ ID NO 33050; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 6304 BP; 1725 A; 1352 C; 1437 G; 1790 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 4; Length 6304;

Best Local Similarity 90.0%; Pred. No. 2.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGCGCAGCGCCCAACAG 20

Db 4208 GATTTCGAGCGCTCAACAG 4189

RESULT 10

ACF67367\_39/c

Continuation (40 of 57) of ACF67367 from base 3900001 (Phototaxidus luminescens nucleoti  
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP Fragment Name Begin End

WP ACF67367\_00 1 110000

WP ACF67367\_01 100001 210000

WP ACF67367\_02 200001 310000

WP ACF67367\_03 300001 410000

WP ACF67367\_04 400001 510000

WP ACF67367\_05 500001 610000

WP ACF67367\_06 600001 710000

WP ACF67367\_07 700001 810000

WP ACF67367\_08 800001 910000

WP ACF67367\_09 900001 1010000

WP ACF67367\_10 1000001 1110000

WP ACF67367\_11 1100001 1210000

WP ACF67367\_12 1200001 1310000

WP ACF67367\_13 1300001 1410000

WP ACF67367\_14 1400001 1510000

WP ACF67367\_15 1500001 1610000

WP ACF67367\_16 1600001 1710000

WP ACF67367\_17 1700001 1810000

WP ACF67367\_18 1800001 1910000

WP ACF67367\_19 1900001 2010000

WP ACF67367\_20 2000001 2110000

WP ACF67367\_21 2100001 2210000

WP ACF67367\_22 2200001 2310000

WP ACF67367\_23 2300001 2410000

WP ACF67367\_24 2400001 2510000

WP ACF67367\_25 2500001 2610000

WP ACF67367\_26 2600001 2710000

WP ACF67367\_27 2700001 2810000

WP ACF67367\_28 2800001 2910000

WP ACF67367\_29 2900001 3010000

WP ACF67367\_30 3000001 3110000

WP ACF67367\_31 3100001 3210000

WP ACF67367\_32 3200001 3310000

WP ACF67367\_33 3300001 3410000

WP ACF67367\_34 3400001 3510000

WP ACF67367\_35 3500001 3610000

WP ACF67367\_36 3600001 3710000

WP ACF67367\_37 3700001 3810000

WP ACF67367\_38 3800001 3910000

WP ACF67367\_39 3900001 4010000

WP ACF67367\_40 4000001 4110000

WP ACF67367\_41 4100001 4210000

WP ACF67367\_42 4200001 4310000

WP ACF67367\_43 4300001 4410000

WP ACF67367\_44 4400001 4510000

WP ACF67367\_45 4500001 4610000

WP ACF67367\_46 4600001 4710000

WP ACF67367\_47 4700001 4810000

WP ACF67367\_48 4800001 4910000

WP ACF67367\_49 4900001 5010000

WP ACF67367\_50 5000001 5110000

WP ACF67367\_51 5100001 5210000

WP ACF67367\_52 5200001 5310000

WP ACF67367\_53 5300001 5410000

WP ACF67367\_54 5400001 5510000

WP ACF67367\_55 5500001 5610000

WP ACF67367\_56 560001 5648894

Query Match 80.0%; Score 16.8; DB 7; Length 110000;  
Best Local Similarity 90.0%; Pred. No. 2.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTGGCAGCGCCCAACAGC 21

Db 69294 ATTGGCAGCGCCCAACAGC 69275

## RESULT 11

ACF65388 08  
Continuation (9 of 13) of ACF65388 from base 800001 (Photorhabdus luminescens nucleotide  
Sequence split into 13 fragments LOCUS ACF65388 Accession ACF65388

WP	Fragment Name	Begin	End
WP	ACF65388_00	1	110000
WP	ACF65388_01	100001	210000
WP	ACF65388_02	200001	310000
WP	ACF65388_03	300001	410000
WP	ACF65388_04	400001	510000
WP	ACF65388_05	500001	610000
WP	ACF65388_06	600001	710000
WP	ACF65388_07	700001	810000
WP	ACF65388_08	800001	910000
WP	ACF65388_09	900001	1010000
WP	ACF65388_10	1000001	1110000
WP	ACF65388_11	1100001	1210000
WP	ACF65388_12	1200001	1225559

Query Match 80.0%; Score 16.8; DB 7; Length 110000;  
Best Local Similarity 90.0%; Pred. No. 2.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTGGCAGCGCCCAACAGC 21

Db 55643 ATTGGCAGCGCCCAACAGC 55662

## RESULT 12

ABN22495  
ID ABN22495 standard; cDNA; 286 BP.

XX AC ABN22495;

XX DT 24-JUN-2002 (first entry)

XX DE Human ORFX polynucleotide sequence SEQ ID NO:13467.

XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
XX KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
XX KW hypertension; hypothyroidism; cholesterol ester storage disease;  
XX KW immune deficiency; immune disorder; infectious disease;  
XX KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
XX KW myasthenia gravis; gene; ss.

XX CS Homo sapiens.

XX FN WO200192523-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US010836.

XX PR 30-MAY-2000; 2000US-0206132P.

XX PR 29-AUG-2000; 2000US-0228716P.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach MD;

XX

DR WPI: 2002-106308/14.  
DR P-PSDB; ABP06743.XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders.

XX PS Disclosure; SEQ ID NO 13467; 1037pp; English.

XX CC The present invention describes substantially purified human proteins  
(referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
in the specification). ABN15762 to ABN27252 encode the human ORFX  
proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
treating or preventing a pathology associated with an ORFX-associated  
disorder in humans, and in the manufacture of a medicament for treating a  
syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
sequences can be used in gene therapy. ORFX sequences can be used in the  
treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
osteoarthritis, neurodegenerative diseases, diabetes mellitus, systemic  
transplantation, cardiovascular diseases, hypothyroidism, cholesterol ester  
storage disease, various immune deficiencies and disorders, infectious  
diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
disease and autoimmune inflammatory eye disease. ORFX proteins are also  
useful for treating burns, incisions, ulcers, for treating osteoporosis,  
bone degenerative disorders, or periodontal disease, and for gut  
protection or regeneration and treatment of lung or liver fibrosis.  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 286 BP; 69 A; 81 C; 74 G; 62 T; 0 U; 0 Other;

Query Match 77.1%; Score 16.2; DB 6; Length 286;  
Best Local Similarity 85.7%; Pred. No. 3.8e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21

Db 47 GTTTCACAGCGCCCAACAGC 67

## RESULT 13

ABZ52448/c

ID ABZ52448 standard; cDNA; 505 BP.

XX AC ABZ52448;

XX DT 28-MAR-2003 (first entry)

XX DE Aspergillus oryzae polynucleotide SEQ ID NO 1561.

XX KW Aspergillus oryzae; fermentation; fungus; industrial; EST;  
XX expressed sequence tag; gene; ss.

XX OS Aspergillus oryzae.

XX PN WO200279476-A1.

XX PD 10-OCT-2002.

XX PF 22-MAR-2002; 2002WO-IB000890.

XX PR 30-MAR-2001; 2001JP-00098371.

XX PA (NAAD-) NAT INST ADVANCED IND SCI &amp; TECHNOLOGY.

XX PA (NARE-) NAT RES INST BREWING.

XX PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;



PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0251475P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-451937/48.  
Isolated polypeptide for treating, preventing and/ or prognosing  
disorders related to the musculoskeletal system including musculoskeletal  
cancers and also for testing and detection e.g. diagnosis.  
Example 2; SEQ ID NO 3087; 781pp + Sequence Listing; English.  
The invention relates to novel genes (AAL34669-AAL37666) and proteins

CC (ABB03087-ABB04109) associated with the musculoskeletal system useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. The genes are isolated from a range of human tissues  
CC disclosed in the specification. The nucleic acids, proteins, antibodies  
CC and (antagonists are useful in the diagnosis, treatment and prevention  
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2847 BP; 614 A; 784 C; 759 G; 690 T; 0 U; 0 Other;  
Query Match 77.1%; Score 16.2; DB 4; Length 2847;  
Best Local Similarity 85.7%; Pred. No. 4.5e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATTGGCGACGGCCCAACAC 21  
Db 882 GATTGGCGACGGCCCAATAGC 862

Search completed: April 29, 2004, 04:57:03  
Job time : 176.453 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 1655.87 Seconds  
(without alignments)  
378.717 Million cell updates/sec

Title: US-10-624-714-6

Perfect score: 21

Sequence: 1 gattgcgcagccccaacagc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estcov:\*

6: em\_estopl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrt:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	297	14	CF945868
2	17.8	84.8	315	14	CF275886
3	17.8	84.8	373	14	CF275956
4	17.8	84.8	380	28	AQ852800

5	17.8	84.8	459	14	CF946066
6	17.8	84.8	498	14	CF946219
7	17.8	84.8	519	14	CF946278
8	17.8	84.8	552	14	CF946096
9	17.8	84.8	571	14	CF946143
10	17.8	84.8	571	14	CF946539
11	17.8	84.8	594	14	CF946224
12	17.8	84.8	594	14	CF946985
13	17.8	84.8	595	14	CF946186
14	17.8	84.8	622	14	CF946427
15	17.8	84.8	622	14	CF946427
16	17.8	84.8	622	14	CF946427
17	17.8	84.8	622	14	CF946427
18	17.8	84.8	622	14	CF946427
19	17.8	84.8	622	14	CF946427
20	17.8	84.8	622	14	CF946427
21	17.8	84.8	622	14	CF946427
22	17.8	84.8	622	14	CF946427
23	17.8	84.8	622	14	CF946427
24	17.8	84.8	622	14	CF946427
25	17.8	84.8	622	14	CF946427
26	17.8	84.8	622	14	CF946427
27	17.8	84.8	622	14	CF946427
28	17.8	84.8	622	14	CF946427
29	17.8	84.8	622	14	CF946427
30	17.8	84.8	622	14	CF946427
31	17.8	84.8	622	14	CF946427
32	17.8	84.8	622	14	CF946427
33	17.8	84.8	622	14	CF946427
34	17.8	84.8	622	14	CF946427
35	17.8	84.8	622	14	CF946427
36	17.8	84.8	622	14	CF946427
37	17.8	84.8	622	14	CF946427
38	17.8	84.8	622	14	CF946427
39	17.8	84.8	622	14	CF946427
40	17.8	84.8	622	14	CF946427
41	17.8	84.8	622	14	CF946427
42	17.8	84.8	622	14	CF946427
43	17.8	84.8	622	14	CF946427
44	17.8	84.8	622	14	CF946427
45	17.8	84.8	622	14	CF946427

#### ALIGNMENTS

RESULT 1	CF945868	CF945868	297 bp	linear	EST 19-NOV-2003
LOCUS	waa92c07.y1	Gregarina niphandrodes	WSU-1 gametocytes	Cdna-3 Library	
DEFINITION	Gregarina niphandrodes	CDNA 5', mRNA sequence.			
ACCESSION	CF945868				
VERSION	CF945868.1	GI:38450686			
KEYWORDS	EST.				
SOURCE	Gregarina niphandrodes				
ORGANISM	Gregarina niphandrodes				
REFERENCE	1 (bases 1 to 297)				
AUTHORS	Cole R., Fogarty S., Tang K., Howe D.K., Sibley L.D., Clifton S., Marra M., Hillier L., Pape D., Martin J., Wylie T., Dante M., Theising B., Bowers X., Gibbons M., Ritter E., Bennett J., Ronko I., Tsagareishvili R., Fedele M., Belaygorod L., Franklin C., Carr L.N., Grow A., Maguire L., Wadkins J., Richey J., Waterston R. and Wilson R.				
TITLE	WashU Neospora Project				
JOURNAL	Unpublished (2003)				
COMMENT	Contact: L. David Sibley WashU Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810				

Email: est@watson.wustl.edu  
 The cDNA library was constructed by Keliang Tang, Washington University, DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 291.  
 Location/Qualifiers

## FEATURES

Source

1..297  
 /organism="Gregarina niphandrodes"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:110365"  
 /dev\_stage="Gametocytes"  
 /lab\_host="GC10 Competent Cells(PGC)"  
 /clone\_lib="Gregarina niphandrodes WSU-1 gametocytes cDNA-3 library"  
 /note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and electroporated into Electrofen Blue cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
 5'GAATTCGGCATTACGGCC(G)n-- insert--  
 GGCCGCTCGGCCACGGATCC3' where n=3-4 G nucleotides."

## ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 297;  
 Best Local Similarity 90.5%; Pred. No. 1.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21  
 |||||  
 Db 107 GATTACGAGCGCCCAACACC 127

RESULT 2  
 LOCUS CF275886 315 bp mRNA linear EST 14-AUG-2003  
 DEFINITION wao03a05.y4 Gregarina niphandrodes WSU-1 gametocytes cDNA library  
 CF275886  
 CF275886.1 GI:33653272

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Gregarina niphandrodes  
 Gregarina niphandrodes  
 Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;  
 Gregarinidae; Gregarina.

## REFERENCE

## AUTHORS

1 (bases 1 to 315)  
 Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,  
 Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T., Dante, M.,  
 Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I.,  
 Tsagarisvili, R., Fedele, M., Belaygorod, L., Franklin, C.,  
 Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.  
 and Wilson, R.

## TITLE

## JOURNAL

## COMMENT

WashU Neospora Project  
 Unpublished (2003)  
 Other ESTs: wao03a05.x3  
 Contact: L. David Sibley  
 WashU Plasmodium EST Project

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 The cDNA library was constructed by Keliang Tang, Washington University, DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 315.  
 Location/Qualifiers

## FEATURES

Source

1..315  
 /organism="Gregarina niphandrodes"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:110365"  
 /dev\_stage="Gametocytes"  
 /lab\_host="Electrofen cells (Stratagene)"  
 /clone\_lib="Gregarina niphandrodes WSU-1 gametocytes cDNA library"  
 /note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and electroporated into Electrofen Blue cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
 5'GAATTCGGCATTACGGCC(G)n-- insert--  
 GGCCGCTCGGCCACGGATCC3' where n=3-4 G nucleotides."

## ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 315;  
 Best Local Similarity 90.5%; Pred. No. 1.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21  
 |||||  
 Db 107 GATTACGAGCGCCCAACACC 127

## RESULT 3

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

CF275956 373 bp mRNA linear EST 14-AUG-2003  
 wao04a05.y4 Gregarina niphandrodes WSU-1 gametocytes cDNA library  
 Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST  
 WALL PROTEIN ; mRNA sequence.  
 CF275956  
 CF275956.1 GI:33653342  
 EST.  
 Gregarina niphandrodes  
 Gregarina niphandrodes  
 Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;  
 Gregarinidae; Gregarina.  
 1 (bases 1 to 373)  
 Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,  
 Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T., Dante, M.,  
 Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I.,  
 Tsagarisvili, R., Fedele, M., Belaygorod, L., Franklin, C.,  
 Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.  
 and Wilson, R.



JOURNAL  
COMMENT

Unpublished (2003)  
Contact: L. David Sibley  
WashU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
The cDNA library was constructed by Keliang Tang, Washington University, DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 408.  
Location/Qualifiers  
1. 459  
/organism="Gregarina niphandrodes"  
/mol\_type="mRNA"  
/db\_xref="taxon:110365"  
/dev\_stage="Gametocytes"  
/lab\_host="GC10 Competent Cells(PCC)"  
/clone\_lib="Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library"  
/note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
5'-GAATTCGGCCATTACGCC(G)n-- insert--  
GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."

FEATURES  
source

1. 459  
/organism="Gregarina niphandrodes"  
/mol\_type="mRNA"  
/db\_xref="taxon:110365"  
/dev\_stage="Gametocytes"  
/lab\_host="GC10 Competent Cells(PCC)"  
/clone\_lib="Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library"  
/note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
5'-GAATTCGGCCATTACGCC(G)n-- insert--  
GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."

ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 459;  
Best Local Similarity 90.5%; Pred. No. 1.6e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21  
|||||  
Db 76 GATTACGAGCGCCCAACACC 96

RESULT 6  
CF946219

LOCUS wab15d03.y1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library  
DEFINITION Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST WALL PROTEIN ;, mRNA sequence.

CF946219 498 bp mRNA linear EST 19-NOV-2003  
CF946219.1 GI:38451037  
EST.

ORGANISM Gregarina niphandrodes  
Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;  
Gregarinidae; Gregarina.

REFERENCE 1 (bases 1 to 498)  
AUTHORS Cole, R., Fogarty, S., Tang, K., Howe, D. K., Sibley, L. D., Clifton, S., Marra, M., Hillier, L., Fape, D., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Nonko, I.,

TITLE  
COMMENT

Tsagarishvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carl, L. M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.  
WashU Neospora Project  
Unpublished (2003)  
Contact: L. David Sibley  
WashU Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
The cDNA library was constructed by Keliang Tang, Washington University, DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University  
Seq primer: -40UP from Gibco  
High quality sequence stop: 498.  
Location/Qualifiers  
1. 498  
/organism="Gregarina niphandrodes"  
/mol\_type="mRNA"  
/db\_xref="taxon:110365"  
/dev\_stage="Gametocytes"  
/lab\_host="GC10 Competent Cells(PCC)"  
/clone\_lib="Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library"  
/note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
5'-GAATTCGGCCATTACGCC(G)n-- insert--  
GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."

FEATURES  
source

1. 498  
/organism="Gregarina niphandrodes"  
/mol\_type="mRNA"  
/db\_xref="taxon:110365"  
/dev\_stage="Gametocytes"  
/lab\_host="GC10 Competent Cells(PCC)"  
/clone\_lib="Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library"  
/note="Vector: Modified pBluescript (pBS SK+); Site\_1: BamHI; Site\_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:  
5'-GAATTCGGCCATTACGCC(G)n-- insert--  
GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."

ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 498;  
Best Local Similarity 90.5%; Pred. No. 1.7e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAGC 21  
|||||  
Db 107 GATTACGAGCGCCCAACACC 127

RESULT 7  
CF946278

LOCUS wab17a05.y1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library  
DEFINITION Gregarina niphandrodes cDNA 5' similar to TR:Q9U732 Q9U732 OOCYST WALL PROTEIN ;, mRNA sequence.

CF946278 519 bp mRNA linear EST 19-NOV-2003  
CF946278.1 GI:38451096  
EST.

ORGANISM Gregarina niphandrodes  
Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;  
Gregarinidae; Gregarina.

```

REFERENCE
AUTHORS
1 (bases 1 to 519)
Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,
Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Dante, M.,
Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Ronko, I.,
Tsagarishvili, R., Fedele, M., Belaygorod, L., Franklin, C.,
Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.
and Wilson, R.

TITLE
JOURNAL
COMMENT
WashU Neospora Project
Unpublished (2003)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The cDNA library was constructed by Keliang Tang, Washington
University. DNA sequencing by: Washington University Genome
Sequencing Center For information on obtaining a clone please
contact: L. David Sibley (sibley@borcim.wustl.edu), Washington
University
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..519
/organism="Gregarina niphandrodes"
/mol_type="mRNA"
/db_xref="taxon:110365"
/lab_host="GC10 Competent Cells (PGC)"
/clone_lib="Gregarina niphandrodes WSU-1 gametocytes
Cdna-3 Library"
/notes="Vector: Modified pBluescript (pBS SK+); Site_1:
BamHI; Site_2: EcoRI; Gametocytes of Gregarina
niphandrodes were obtained from Tenebrio molitor by Dr.
Charlotte Omoto (Washington State University). The cDNA
library was constructed by Keliang Tang, and Robert Cole
at Washington University. cDNA was synthesized from
poly(A)+ mRNA using the template-switching PCR method
(SMART cDNA Kit, BD Biosciences). First strand cDNA was
reverse transcribed using the CDS III/3' primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector (obtained from Michael White, Montana
State University) containing directional SfiI sites, and
electroporated into Electroten Blue cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'GAATTGGCCATTACGGCC(G)n-- insert--
GGCGGCTCGGCCCGCCACGATCC3'where n=3-4 G nucleotides."

FEATURES
source
1..519
/organism="Gregarina niphandrodes"
/mol_type="mRNA"
/db_xref="taxon:110365"
/lab_host="GC10 Competent Cells (PGC)"
/clone_lib="Gregarina niphandrodes WSU-1 gametocytes
Cdna-3 Library"
/notes="Vector: Modified pBluescript (pBS SK+); Site_1:
BamHI; Site_2: EcoRI; Gametocytes of Gregarina
niphandrodes were obtained from Tenebrio molitor by Dr.
Charlotte Omoto (Washington State University). The cDNA
library was constructed by Keliang Tang, and Robert Cole
at Washington University. cDNA was synthesized from
poly(A)+ mRNA using the template-switching PCR method
(SMART cDNA Kit, BD Biosciences). First strand cDNA was
reverse transcribed using the CDS III/3' primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector (obtained from Michael White, Montana
State University) containing directional SfiI sites, and
electroporated into Electroten Blue cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'GAATTGGCCATTACGGCC(G)n-- insert--
GGCGGCTCGGCCCGCCACGATCC3'where n=3-4 G nucleotides."

ORIGIN
Query Match 84.8%; Score 17.8; DB 14; Length 519;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGGCAGCGCCCAACAGC 21
|||||
Db 107 GATTACGACGCGCCCAACACC 127

RESULT 8
CF946096 552 bp mRNA linear EST 19-NOV-2003
LOCUS
DEFINITION
wab05a01.y1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library
Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST
WALL PROTEIN ;, mRNA sequence.
ACCESSION
CF946096
VERSION
CF946096.1 GI:38450914
KEYWORDS
EST.
SOURCE
Gregarina niphandrodes

ORGANISM
Gregarina niphandrodes
Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;
Gregarinidae; Gregarina.
1 (bases 1 to 552)
Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,
Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Dante, M.,
Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Ronko, I.,
Tsagarishvili, R., Fedele, M., Belaygorod, L., Franklin, C.,
Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.
and Wilson, R.
WashU Neospora Project
Unpublished (2003)
Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The cDNA library was constructed by Keliang Tang, Washington
University. DNA sequencing by: Washington University Genome
Sequencing Center For information on obtaining a clone please
contact: L. David Sibley (sibley@borcim.wustl.edu), Washington
University
Seq primer: -40UP from Gibco
High quality sequence stop: 552.
Location/Qualifiers
1..552
/organism="Gregarina niphandrodes"
/mol_type="mRNA"
/db_xref="taxon:110365"
/lab_host="GC10 Competent Cells (PGC)"
/clone_lib="Gregarina niphandrodes WSU-1 gametocytes
Cdna-3 Library"
/notes="Vector: Modified pBluescript (pBS SK+); Site_1:
BamHI; Site_2: EcoRI; Gametocytes of Gregarina
niphandrodes were obtained from Tenebrio molitor by Dr.
Charlotte Omoto (Washington State University). The cDNA
library was constructed by Keliang Tang, and Robert Cole
at Washington University. cDNA was synthesized from
poly(A)+ mRNA using the template-switching PCR method
(SMART cDNA Kit, BD Biosciences). First strand cDNA was
reverse transcribed using the CDS III/3' primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector (obtained from Michael White, Montana
State University) containing directional SfiI sites, and
electroporated into Electroten Blue cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'GAATTGGCCATTACGGCC(G)n-- insert--
GGCGGCTCGGCCCGCCACGATCC3'where n=3-4 G nucleotides."

ORIGIN
Query Match 84.8%; Score 17.8; DB 14; Length 552;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTGGCAGCGCCCAACAGC 21
|||||
Db 165 GATTACGACGCGCCCAACACC 185

RESULT 9
CF946143 571 bp mRNA linear EST 19-NOV-2003
LOCUS
DEFINITION
wab08g02.y1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library
Gregarina niphandrodes cDNA 5' similar to TR:Q9U734 Q9U734 OOCYST
WALL PROTEIN ;, mRNA sequence.

```

ACCESSION	CF946143	LOCUS	CF946539	571 bp	mRNA	linear	EST 19-NOV-2003
VERSION	CF946143.1	DEFINITION	wab25b12.v1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library				
KEYWORDS	EST.		Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST				
SOURCE	Gregarina niphandrodes	ACCESSION	CF946539				
ORGANISM	Gregarina niphandrodes	VERSION	CF946539.1	GI:38451357			
	Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;	KEYWORDS	EST.				
REFERENCE	1 (bases 1 to 571)	SOURCE	Gregarina niphandrodes				
AUTHORS	Cole, R., Fogarty, S., Tang, K., Howe, D. K., Sibley, L. D., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I., Tsagarisshvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carr, L. M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.	ORGANISM	Gregarina niphandrodes				
	WashU Neospora Project		Gregarina niphandrodes				
TITLE	Unpublished (2003)		Gregarina niphandrodes				
JOURNAL	WashU Plasmodium EST Project		Gregarina niphandrodes				
COMMENT	Contact: L. David Sibley Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu		Gregarina niphandrodes				
	The cDNA library was constructed by Keliang Tang, Washington University. DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University		Gregarina niphandrodes				
FEATURES	Seq primer: -40UP from Gibco		Gregarina niphandrodes				
source	High quality sequence stop: 495.		Gregarina niphandrodes				
	Location/Qualifiers		Gregarina niphandrodes				
	1..571		Gregarina niphandrodes				
	/organism="Gregarina niphandrodes"		Gregarina niphandrodes				
	/mol_type="mRNA"		Gregarina niphandrodes				
	/db_xref="taxon:110365"		Gregarina niphandrodes				
	/dev_stage="Gametocytes"		Gregarina niphandrodes				
	/lab_hosts="GC10 Competent Cells (PGC)"		Gregarina niphandrodes				
	/clone_lib="Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library"		Gregarina niphandrodes				
	/note="Vector: Modified pBluescript (pBS SK+); Site_1: BamHI; Site_2: EcoRI; Gametocytes of Gregarina niphandrodes were obtained from Tenebrio molitor by Dr. Charlotte Omoto (Washington State University). The cDNA library was constructed by Keliang Tang, and Robert Cole at Washington University. cDNA was synthesized from poly(A)+ mRNA using the template-switching PCR method (SMART cDNA Kit, BD Biosciences). First strand cDNA was reverse transcribed using the CDS III/3' primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector (obtained from Michael White, Montana State University) containing directional SfiI sites, and electroporated into Electrofen Blue cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence: 5'GAATTCGGCGCATTCAGCC(G)n-- insert-- GGCGCGCTCGGCCACGATCC3'where n=3-4 G nucleotides."		Gregarina niphandrodes				
ORIGIN			Gregarina niphandrodes				
Query Match	84.8%; Score 17.8; DB 14; Length 571;		Gregarina niphandrodes				
Best Local Similarity	90.5%; Pred. No. 1.8e+03;		Gregarina niphandrodes				
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		Gregarina niphandrodes				
QY	1 GATTGGCGAGCGCCCAACAGC 21		Gregarina niphandrodes				
Db	164 GATTACGAGCGCCCAACACC 184		Gregarina niphandrodes				
RESULT 10			Gregarina niphandrodes				
CF946539			Gregarina niphandrodes				

```

RESULT 11
CF946224
LOCUS
DEFINITION
  CF946224.1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library
  Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST
  WALL PROTEIN 1, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
  CF946224.1 GI:38451042
ORGANISM
  Gregarina niphandrodes
  Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;
  Gregarinidae; Gregarina.
  1 (bases 1 to 594)
REFERENCE
  Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,
  Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Dante, M.,
  Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I.,
  Tsagareishvili, R., Fedele, M., Belaygorod, L., Franklin, C.,
  Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.
  and Wilson, R.
  WashU Neospora Project
  Unpublished (2003)
  Contact: L. David Sibley
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286-1800
  Fax: 314 286 1810
  Email: estowatson.wustl.edu
  The cDNA library was constructed by Keliang Tang, Washington
  University. DNA sequencing by: Washington University Genome
  Sequencing Center For information on obtaining a clone please
  contact: L. David Sibley (sibley@borcim.wustl.edu), Washington
  University.
  Seq primer: -40UP from Gibco.
  Location/Qualifiers
    1..594
      /organism="Gregarina niphandrodes"
      /mol_type="mRNA"
      /db_xref="taxon:110365"
      /dev_stage="Gametocytes"
      /lab_host="GC10 Competent Cells (PGC)"
      /clone_lib="Gregarina niphandrodes WSU-1 gametocytes
      Cdna-3 Library"
      /note="Vector: Modified pBluescript (pBS SK+); Site 1:
      BamHI; Site 2: EcoRI; Gametocytes of Gregarina
      niphandrodes were obtained from Tenebrio molitor by Dr.
      Charlotte Omoto (Washington State University). The cDNA
      library was constructed by Keliang Tang, and Robert Cole
      at Washington University. cDNA was synthesized from
      poly(A)+ mRNA using the template-switching PCR method
      (SMART cDNA Kit, BD Biosciences). First strand cDNA was
      reverse transcribed using the CDS III/3' primer and a 5'
      template switch primer (Smart IV primer). The product of
      the first strand synthesis was PCR amplified using the
      same primer set and the fragments were digested with SfiI.
      The fragments were size selected, ligated into a modified
      pBluescript vector (obtained from Michael White, Montana
      State University) containing directional SfiI sites, and
      electroporated into ElectroTen Blue cells. Vector: SfiI
      sites were added to the multiple cloning region of
      pBluescript SK+ between the BamHI/EcoRI sites. The
      modified polylinker has the following sequence:
      5'GAATTCGGCCATTACGCC(G)n-- insert--
      GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."
ORIGIN
  Query Match      84.8%; Score 17.8; DB 14; Length 594;
  Best Local Similarity 90.5%; Pred. No. 1.8e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  Qy 1 GATTGCGCAGCGCCCAACAGC 21

RESULT 12
CF946985
LOCUS
DEFINITION
  CF946985.1 Gregarina niphandrodes WSU-1 gametocytes Cdna-3 Library
  Gregarina niphandrodes cDNA 5' similar to TR:Q9U733 Q9U733 OOCYST
  WALL PROTEIN 1, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
  CF946985.1 GI:38451803
ORGANISM
  Gregarina niphandrodes
  Eukaryota; Alveolata; Apicomplexa; Gregarina; Eugregarinida;
  Gregarinidae; Gregarina.
  1 (bases 1 to 594)
REFERENCE
  Cole, R., Fogarty, S., Tang, K., Howe, D.K., Sibley, L.D., Clifton, S.,
  Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Dante, M.,
  Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I.,
  Tsagareishvili, R., Fedele, M., Belaygorod, L., Franklin, C.,
  Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R.
  and Wilson, R.
  WashU Neospora Project
  Unpublished (2003)
  Contact: L. David Sibley
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286-1800
  Fax: 314 286 1810
  Email: estowatson.wustl.edu
  The cDNA library was constructed by Keliang Tang, Washington
  University. DNA sequencing by: Washington University Genome
  Sequencing Center For information on obtaining a clone please
  contact: L. David Sibley (sibley@borcim.wustl.edu), Washington
  University.
  Seq primer: -40UP from Gibco.
  Location/Qualifiers
    1..594
      /organism="Gregarina niphandrodes"
      /mol_type="mRNA"
      /db_xref="taxon:110365"
      /dev_stage="Gametocytes"
      /lab_host="GC10 Competent Cells (PGC)"
      /clone_lib="Gregarina niphandrodes WSU-1 gametocytes
      Cdna-3 Library"
      /note="Vector: Modified pBluescript (pBS SK+); Site 1:
      BamHI; Site 2: EcoRI; Gametocytes of Gregarina
      niphandrodes were obtained from Tenebrio molitor by Dr.
      Charlotte Omoto (Washington State University). The cDNA
      library was constructed by Keliang Tang, and Robert Cole
      at Washington University. cDNA was synthesized from
      poly(A)+ mRNA using the template-switching PCR method
      (SMART cDNA Kit, BD Biosciences). First strand cDNA was
      reverse transcribed using the CDS III/3' primer and a 5'
      template switch primer (Smart IV primer). The product of
      the first strand synthesis was PCR amplified using the
      same primer set and the fragments were digested with SfiI.
      The fragments were size selected, ligated into a modified
      pBluescript vector (obtained from Michael White, Montana
      State University) containing directional SfiI sites, and
      electroporated into ElectroTen Blue cells. Vector: SfiI
      sites were added to the multiple cloning region of
      pBluescript SK+ between the BamHI/EcoRI sites. The
      modified polylinker has the following sequence:
      5'GAATTCGGCCATTACGCC(G)n-- insert--
      GGCCGCTCGGCCACGATCC3' where n=3-4 G nucleotides."
ORIGIN
  Query Match      84.8%; Score 17.8; DB 14; Length 594;
  Best Local Similarity 90.5%; Pred. No. 1.8e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  Qy 1 GATTGCGCAGCGCCCAACAGC 21

```







5'GAATTCGCGCCATTACGGCC(G)n-- insert--  
GGCGCGCTCGGCCACGATGCC'where n=3-4 G nucleotides."

ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 622;  
Best Local Similarity 90.5%; Pred. No. 1.8e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTCGCGCGCCCAACAGC 21  
|||||  
Db 164 GATTTCGCGCGCCCAACACC 184

BB838943 467 bp mRNA linear EST 21-NOV-2001  
BB838943 RIKEN full-length enriched, 8 cells embryo Mus musculus  
cDNA clone B860006D20 5', mRNA sequence.

ACCESSION BB838943  
VERSION BB838943.1 GI:17039674  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (Bases 1 to 467)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,  
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,  
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,  
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.  
and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

/clone="B860006D20"  
/cell type="8 cells"  
/dev stage="8 cells embryo"  
/clone\_lib="RIKEN full-length enriched, 8 cells embryo"

ORIGIN

Query Match 81.0%; Score 17; DB 10; Length 467;  
Best Local Similarity 100.0%; Pred. No. 3.7e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TGGCGAGCGGCCCAACAG 20  
|||||  
Db 113 TGGCGAGCGGCCCAACAG 97

Search completed: April 29, 2004, 11:37:01  
Job time : 1659.87 secs

FEATURES  
source  
1. .467  
Location/Qualifiers  
/organism="Mus musculus"  
/mol type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="taxon:10090"

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 37.0588 Seconds  
(without alignments)  
314.472 Million cell updates/sec

Title: US-10-624-714-6

Perfect score: 21

Sequence: 1 gattgcgagcgcccaacagc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21	100.0	4403765	3	US-09-103-840A-2
C 2	21	100.0	4411529	3	US-09-103-840A-1
C 3	16.8	80.0	546	4	US-09-252-991A-11071
C 4	16.8	80.0	909	4	US-09-252-991A-11185
C 5	15.4	73.3	190	2	US-08-890-980-18
C 6	15.4	73.3	190	3	US-08-890-979-18
C 7	15.4	73.3	190	3	US-09-032-894-18
C 8	15.4	73.3	190	3	US-09-031-626-18
C 9	15.4	73.3	1002	2	US-08-890-980-5
C 10	15.4	73.3	1002	3	US-08-890-979-5
C 11	15.4	73.3	1002	3	US-09-032-894-5
C 12	15.4	73.3	1002	3	US-09-032-894-95
C 13	15.4	73.3	1002	3	US-09-031-626-5
C 14	15.4	73.3	1002	3	US-09-031-626-95
C 15	15.4	73.3	1113	4	US-09-489-039A-1851
C 16	15.4	73.3	1494	4	US-09-252-991A-14961
C 17	15.4	73.3	1509	4	US-09-252-991A-14669
C 18	15.4	73.3	1668	4	US-09-252-991A-15085
C 19	15.4	73.3	2220	4	US-09-252-991A-16207
C 20	15.2	72.4	363	4	US-09-252-991A-15666
C 21	15.2	72.4	531	4	US-09-252-991A-15586
C 22	15.2	72.4	888	4	US-09-489-039A-804
C 23	15.2	72.4	981	4	US-09-252-991A-15726
C 24	15.2	72.4	1209	4	US-09-543-681A-1103
C 25	15.2	72.4	1455	3	US-09-276-531-33
C 26	15.2	72.4	1476	4	US-09-489-039A-665
C 27	15.2	72.4	1551	4	US-09-489-039A-489

C 28	15.2	72.4	26173	4	US-09-453-702B-69
C 29	15.2	72.4	31880	4	US-09-453-702B-242
C 30	15.2	72.4	38155	4	US-09-453-702B-79
C 31	14.8	70.5	282	4	US-09-489-039A-2170
C 32	14.8	70.5	369	4	US-09-489-039A-2134
C 33	14.8	70.5	447	4	US-09-489-039A-2143
C 34	14.8	70.5	492	4	US-09-540-236-727
C 35	14.8	70.5	693	4	US-09-252-991A-825
C 36	14.8	70.5	791	3	US-08-998-416-468
C 37	14.8	70.5	1650	4	US-09-489-039A-6815
C 38	14.8	70.5	2212	4	US-09-071-035-459
C 39	14.8	70.5	6165	4	US-09-134-000C-3207
C 40	14.8	70.5	6168	4	US-09-071-035-457
C 41	14.8	70.5	6168	4	US-09-071-035-461
C 42	14.8	70.5	6168	4	US-09-071-035-465
C 43	14.8	70.5	4403765	3	US-09-103-840A-2
C 44	14.8	70.5	4411529	3	US-09-103-840A-1
C 45	14.6	69.5	555	4	US-09-634-238-185

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007-00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Query Match 100.0%; Score 21; DB 3; Length 4403765;

Best Local Similarity 100.0%; Pred. No. 0.54;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 GATTGCGAGCGCCCAACAGC 21

|||||

Db 472886 GATTGCGAGCGCCCAACAGC 472866

RESULT 2

US-09-103-840A-1/c

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007-00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 21; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.54; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

Qy 1 GATTGGCAGCGCCCAACAGC 21
Db 472802 GATTGGCAGCGCCCAACAGC 472782

RESULT 3
US-09-252-991A-11071
; Sequence 11071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11071
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11071

Query Match      80.0%; Score 16.8; DB 4; Length 546;
Best Local Similarity 90.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ATTGGCAGCGCCCAACAGC 21
Db 297 ATTGGCAGCGCCCAACAGC 316

RESULT 4
US-09-252-991A-11185/c
; Sequence 11185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11185
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11185

Query Match      80.0%; Score 16.8; DB 4; Length 909;
Best Local Similarity 90.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 ATTGGCAGCGCCCAACAGC 21
Db 652 ATTGGCAGCGCCCAACAGC 633

RESULT 5
US-08-890-980-18/c
; Sequence 18, Application US/08890980
; Patent No. 5998141
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: SR-B1 NUCLEIC ACIDS AND USES THEREFOR
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/890,980
; FILING DATE: 10-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Attold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MIA-005.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-890-980-18

Query Match      73.3%; Score 15.4; DB 2; Length 190;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GCGCAGCGCCCAACAGC 21
Db 58 GCCCAGCGCCCAACAGC 42

RESULT 6
US-08-890-979-18/c
; Sequence 18, Application US/08890979
; Patent No. 6030778
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Ordovas, Jose M.
; TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
```



: GENERAL INFORMATION:  
 : APPLICANT: Acton, Susan L.  
 : APPLICANT: Ordovas, Jose M.  
 : TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS  
 : TITLE OF INVENTION: DISORDERS  
 : NUMBER OF SEQUENCES: 75  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 : STREET: One Post Office Square  
 : CITY: Boston  
 : STATE: MA  
 : COUNTRY: USA  
 : ZIP: 02109-2170  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent In Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/890,979  
 : FILING DATE: 10-JUL-1997  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Arnold, Beth E.  
 : REGISTRATION NUMBER: 35,430  
 : REFERENCE/DOCKET NUMBER: MIA-005.02  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 617-832-1000  
 : TELEFAX: 617-832-7000  
 : INFORMATION FOR SEQ ID NO: 5:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1002 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: DNA (genomic)  
 : US-08-890-979-5

```
Query Match      73.3%   Score 15.4; DB 3; Length 1002;
Best Local Similarity 94.1%; Pred.No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

RESULT 11
US-09-032-894-5/c
; Sequence 5, Application US/09032894
; Patent No. 6130041
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: SR-BI NUCLEIC
; FILE REFERENCE: MIA-005.03
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1998-02-27
; EARLIER APPLICATION NUMBER: 08/890
; EARLIER FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Human.
US-09-032-894-5

```

```

Query Match          73.3%; Score 15.4; DB 3; Length 1002;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0

Qy      5 GCGCAGCGCCCAACAGC 21
      |||

```

```

Db      870 GCCCAGCGCCCAACAGC 854

RESULT 12
US-09-032-894-95/c
; Sequence 95, Application US/09032894
; Patent No. 6130041
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: SR-BI NUCLEIC ACIDS AND USES THEREFOR
; FILE REFERENCE: WIA-005-03
; CURRENT APPLICATION NUMBER: US/09/032,894
; CURRENT FILING DATE: 1998-02-27
; EARLIER APPLICATION NUMBER: 08/890,980
; EARLIER FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 95
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Human
US-09-032-894-95

Query Match          73.3%; Score 15.4; DB 3; Length 1002;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0;

Qy      5 GCGCAGCGCCCAACAGC 21
       |||||
Db      870 GCCCAGCGCCCAACAGC 854

```

```

RESULT 13
US-09-031-626-5/c
; Sequence 5, Application US/09031626
; Patent No. 6228581
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Ordovas, Jose M.
; TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS
; FILE REFERENCE: MIA-005.04
; CURRENT APPLICATION NUMBER: US/09/031,626
; CURRENT FILING DATE: 1998-02-27
; EARLIER APPLICATION NUMBER: 08/890,979
; EARLIER FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Human
US-09-031-626-5

```

Query Match	73.3%	Score 15.4;	DB 3;	Length 1002;
Best Local Similarity	94.1%;	Pred. No. 1.4e-02;		
Matches 16;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	5	GCGCAGCGCCCAACAGC	21	
Db	870	GCCCAGCGCCCAACAGC	854	

RESULT 14  
US-09-031-626-95/c  
; Sequence 95, Application US/09031626  
; Patent No. 6228581  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Ordovas, Jose M.  
; TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS  
; FILE REFERENCE: MIA-005.04

; CURRENT APPLICATION NUMBER: US/09/031,626  
 ; CURRENT FILING DATE: 1998-02-27  
 ; EARLIER APPLICATION NUMBER: 08/890,979  
 ; EARLIER FILING DATE: 1997-07-10  
 ; NUMBER OF SEQ ID NOS: 121  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 95  
 ; LENGTH: 1002  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ;  
 US-09-031-626-95

Query Match 73.3%; Score 15.4; DB 3; Length 1002;  
 Best Local Similarity 94.1%; Pred. No. 1.4e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GCCGAGCGCCCAACAGC 21  
 |||||  
 Db 870 GCCGAGCGCCCAACAGC 854

RESULT 15

US-09-489-039A-1851/C  
 ; Sequence 1851, Application US/09/489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709,2034001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 1851  
 ; LENGTH: 1113  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 ;  
 US-09-489-039A-1851

Query Match 73.3%; Score 15.4; DB 4; Length 1113;  
 Best Local Similarity 94.1%; Pred. No. 1.4e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GCCGAGCGCCCAACAGC 21  
 |||||  
 Db 184 GCCGAGCGCCCAACAGC 168

Search completed: April 29, 2004, 11:46:01  
 Job time : 53.0588 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 392 Seconds  
(without alignments)  
242.328 Million cell updates/sec

Title: US-10-624-714-6

Perfect score: 21

Sequence: 1 gattcgagcgcccaacgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:  
14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:  
19: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	16.8	80.0	2423	9	US-09-925-297-23
C 2	16.4	78.1	990	16	US-10-369-493-39445
C 3	16.4	78.1	990	16	US-10-369-493-39829
C 4	16.4	78.1	1038	16	US-10-369-493-39077
C 5	16.2	77.1	1488	13	US-10-282-122A-12949
C 6	16.2	77.1	2361	13	US-10-403-571-3
C 7	16.2	77.1	2847	9	US-09-764-877-3086
C 8	16.2	77.1	2847	9	US-09-764-877-3087
C 9	16.2	77.1	2847	16	US-10-242-515-3086
C 10	16.2	77.1	2847	16	US-10-242-515-3087
C 11	16	76.2	9025608	15	US-10-156-761-1
C 12	15.8	75.2	763	9	US-09-910-943-24
C 13	15.4	73.3	190	10	US-09-779-152-18
C 14	15.4	73.3	190	15	US-10-023-610-18

C 15	15.4	73.3	190	17	US-10-212-848-18
C 16	15.4	73.3	822	9	US-09-815-242-5920
C 17	15.4	73.3	822	13	US-10-282-122A-20217
C 18	15.4	73.3	822	16	US-10-369-493-24377
C 19	15.4	73.3	1002	10	US-09-779-152-5
C 20	15.4	73.3	1002	10	US-09-779-152-95
C 21	15.4	73.3	1002	15	US-10-023-610-5
C 22	15.4	73.3	1002	15	US-10-023-610-95
C 23	15.4	73.3	1002	17	US-10-212-848-5
C 24	15.4	73.3	1002	17	US-10-212-848-95
C 25	15.4	73.3	1254	9	US-09-738-626-2084
C 26	15.4	73.3	1311	13	US-10-282-122A-39976
C 27	15.4	73.3	1311	13	US-10-282-122A-39715
C 28	15.4	73.3	1842	13	US-10-282-122A-40760
C 29	15.4	73.3	3762	13	US-10-282-122A-33752
C 30	15.4	73.3	3933	16	US-10-369-493-32532
C 31	15.4	73.3	8867	10	US-09-532-367A-22
C 32	15.4	73.3	69081	13	US-10-087-192-1192
C 33	15.4	73.3	3309400	9	US-09-738-626-1
C 34	15.2	72.4	199	13	US-10-424-599-30006
C 35	15.2	72.4	209	9	US-09-783-590-4878
C 36	15.2	72.4	388	9	US-09-960-352-13712
C 37	15.2	72.4	493	15	US-10-029-386-24963
C 38	15.2	72.4	500	15	US-10-029-386-22919
C 39	15.2	72.4	560	10	US-09-918-995-9261
C 40	15.2	72.4	752	13	US-10-424-599-62175
C 41	15.2	72.4	975	16	US-10-355-956-1
C 42	15.2	72.4	1338	16	US-10-369-493-40990
C 43	15.2	72.4	1473	10	US-09-796-753-47
C 44	15.2	72.4	1473	15	US-10-156-761-2639
C 45	15.2	72.4	1481	15	US-10-198-846-11017

#### ALIGNMENTS

#### RESULT 1

US-09-925-297-23/c  
; Sequence 23, Application US/09925297  
; Patent No. US2002081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: P4105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 23  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (4)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (12)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (18)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (34)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature

Sequence 18, Appl  
Sequence 5920, Ap  
Sequence 20217, A  
Sequence 24377, A  
Sequence 5, Appl  
Sequence 95, Appl  
Sequence 5, Appl  
Sequence 95, Appl  
Sequence 5, Appl  
Sequence 95, Appl  
Sequence 2084, Ap  
Sequence 38976, A  
Sequence 39715, A  
Sequence 40760, A  
Sequence 33752, A  
Sequence 32532, A  
Sequence 22, Appl  
Sequence 1192, Ap  
Sequence 1, Appl  
Sequence 30006, A  
Sequence 4878, Ap  
Sequence 13712, A  
Sequence 24963, A  
Sequence 22919, A  
Sequence 9261, Ap  
Sequence 62175, A  
Sequence 1, Appl  
Sequence 40990, A  
Sequence 47, Appl  
Sequence 2639, Ap  
Sequence 11017, A

LOCATION: (2409)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (2422)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-23

Query Match 80.0%; Score 16.8; DB 9; Length 2423;  
Best Local Similarity 90.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAG 20  
|||||  
DB 1005 GATTGGCAGCGCCCAACAG 986

## RESULT 2

US-10-369-493-39445/c  
Sequence 39445, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39445  
LENGTH: 990  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39445

Query Match 78.1%; Score 16.4; DB 16; Length 990;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGCGCCCAACAGC 21  
|||||  
DB 842 TGGCAGCGCCCAACAGC 825

## RESULT 3

US-10-369-493-39829/c  
Sequence 39829, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39829  
LENGTH: 990  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39829

Query Match 78.1%; Score 16.4; DB 16; Length 990;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGCGCCCAACAGC 21  
|||||  
DB 842 TGGCAGCGCCCAACAGC 825

## RESULT 4

US-10-369-493-39077/c  
Sequence 39077, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39077  
LENGTH: 1038  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39077

Query Match 78.1%; Score 16.4; DB 16; Length 1038;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGCGCCCAACAGC 21  
|||||  
DB 890 TGGCAGCGCCCAACAGC 873

## RESULT 5

US-10-282-122A-12949/c  
Sequence 12949, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09

```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12949
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-12949

Query Match          77.1%; Score 16.2; DB 13; Length 1488;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTGGCGCAGCGCCCAACAGC 21
Db      560 GCTTGGCGGCGGCTCAACAGC 540

RESULT 6
US-10-403-571-3
; Sequence 3, Application US/10403571
; Publication No. US2004006763A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Nancy
; APPLICANT: Golling, Gregory
; APPLICANT: Amsterdam, Adam
; APPLICANT: Sun, Zhaoxia
; TITLE OF INVENTION: Developmental Mutations in Zebrafish
; FILE REFERENCE: 01997/539002
; CURRENT APPLICATION NUMBER: US/10/403,571
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 60/369,760
; PRIOR FILING DATE: 2002-03-29
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2361
; TYPE: DNA
; ORGANISM: Danio rerio
US-10-403-571-3

Query Match          77.1%; Score 16.2; DB 13; Length 2361;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTGGCGCAGCGCCCAACAGC 21
Db      207 GAGAGCGCAGAGCCCAACAGC 227

RESULT 7
US-09-764-877-3086/c
; Sequence 3086, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 3086
; LENGTH: 2847
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3086

Query Match          77.1%; Score 16.2; DB 9; Length 2847;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTGGCGCAGCGCCCAACAGC 21
Db      882 GATGGCGCAGCGGCCAATAGC 862

RESULT 8
US-09-764-877-3087/c
; Sequence 3087, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3087
; LENGTH: 2847
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3087

Query Match          77.1%; Score 16.2; DB 9; Length 2847;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATTGGCGCAGCGCCCAACAGC 21
Db      882 GATGGCGCAGCGGCCAATAGC 862

RESULT 9
US-10-242-515-3086/c
; Sequence 3086, Application US/10242515
; Publication No. US20040009488A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
```



; TITLE OF INVENTION: DIAGNOSTIC ASSAYS AND KITS FOR BODY MASS AND  
; FILE REFERENCE: MNI-172CP2  
; CURRENT APPLICATION NUMBER: US/09/779,152  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 08/890,979  
; PRIOR FILING DATE: 1997-07-10  
; NUMBER OF SEQ ID NOS: 121  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 190  
; TYPE: DNA  
; ORGANISM: Human  
; US-09-779-152-18

Query Match 73.3%; Score 15.4; DB 10; Length 190;  
Best Local Similarity 94.1%; Pred. No. 4.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 5 GCGCAGCGCCCAACAGC 21  
Db 58 GCCCAGCGCCCAACAGC 42

RESULT 14

US-10-023-610-18/c  
; Sequence 18, Application US/10023610  
; Publication No. US20030023059A1  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; TITLE OF INVENTION: SR-BI NUCLEIC ACIDS AND USES THEREFOR  
; FILE REFERENCE: MIA-005.03  
; CURRENT APPLICATION NUMBER: US/10/023,610  
; CURRENT FILING DATE: 2001-12-17  
; EARLIER APPLICATION NUMBER: 09/686,106  
; EARLIER FILING DATE: 2000-10-10  
; EARLIER APPLICATION NUMBER: 09/032,894  
; EARLIER FILING DATE: 1998-02-27  
; EARLIER APPLICATION NUMBER: 08/890,980  
; EARLIER FILING DATE: 1997-07-10  
; NUMBER OF SEQ ID NOS: 121  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 190  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-023-610-18

Query Match 73.3%; Score 15.4; DB 15; Length 190;  
Best Local Similarity 94.1%; Pred. No. 4.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 5 GCGCAGCGCCCAACAGC 21  
Db 58 GCCCAGCGCCCAACAGC 42

RESULT 15

US-10-212-848-18/c  
; Sequence 18, Application US/10212848  
; Publication No. US20040023225A1  
; GENERAL INFORMATION:  
; APPLICANT: MCCARTHY, Jeanette  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING RISK FACTORS  
; TITLE OF INVENTION: FOR ABNORMAL LIPID LEVELS AND THE DISEASES AND DISORDERS  
; TITLE OF INVENTION: ASSOCIATED THEREWITH  
; FILE REFERENCE: MNI-012  
; CURRENT APPLICATION NUMBER: US/10/212,848  
; CURRENT FILING DATE: 2002-08-05  
; NUMBER OF SEQ ID NOS: 121  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 190

; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-212-848-18  
Query Match 73.3%; Score 15.4; DB 17; Length 190;  
Best Local Similarity 94.1%; Pred. No. 4.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 5 GCGCAGCGCCCAACAGC 21  
Db 58 GCCCAGCGCCCAACAGC 42  
Search completed: April 29, 2004, 20:44:30  
Job time : 404 secs

This Page Blank (uspio)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 316.894 Seconds  
(without alignments)  
2872.264 Million cell updates/sec

Title: US-10-624-714-6

Perfect score: 21  
Sequence: 1 gattgcgcagcccaacagc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	21	100.0	15178	1	AE006945	AE006945 Mycobacte
C 2	21	100.0	324050	1	BX248335	BX248335 Mycobacte
C 3	21	100.0	342416	1	BX842573	BX842573 Mycobacte
C 4	18.4	87.6	412	6	AX071496	AX071496 Sequence
C 5	17.8	84.8	7005	8	SCU22049	U22049 Schizophyll
C 6	17.8	84.8	110000	8	AC125735_2	Continuation (3 of
C 7	17.8	84.8	255490	1	AF005331	AF005331 Vibrio vu
C 8	17.8	84.8	301235	1	AE016799	AE016799 Vibrio vu
C 9	17	81.0	72580	8	NCB11H24	AL670005 Neurospor
C 10	16.8	80.0	2196	4	AB008852	AB008852 Bos tauru
C 11	16.8	80.0	2809	3	AF059613	AF059613 Drosophil
C 12	16.8	80.0	3377	3	BT003191	BT003191 Drosophil
C 13	16.8	80.0	11259	1	AE004465	AE004465 Pseudomon
C 14	16.8	80.0	17665	1	AF521878	AF521878 Streptomy
C 15	16.8	80.0	21788	2	AC013151	AC013151 Drosophil
C 16	16.8	80.0	51362	2	AC014827	AC014827 Drosophil
C 17	16.8	80.0	91809	2	AC012832	AC012832 Drosophil
C 18	16.8	80.0	126322	2	AC008095	AC008095 Drosophil
C 19	16.8	80.0	158402	3	AC008343	AC008343 Drosophil
C 20	16.8	80.0	163072	3	AC009032	AC009032 Drosophil
C 21	16.8	80.0	170994	3	AC007520	AC007520 Drosophil
C 22	16.8	80.0	174551	3	AC010715	AC010715 Drosophil
C 23	16.8	80.0	196672	3	AC004767	AC004767 Drosophil
C 24	16.8	80.0	223577	2	AC135374	AC135374 Rattus no
C 25	16.8	80.0	233210	2	AC097693	AC097693 Rattus no
C 26	16.8	80.0	260337	3	AE003804	AE003804 Drosophil
C 27	16.8	80.0	265187	3	AE003563	AE003563 Drosophil
C 28	16.8	80.0	271178	3	AE003810	AE003810 Drosophil
C 29	16.8	80.0	349907	1	BX571874	BX571874 Phototrab
C 30	16.4	78.1	10029	1	AE013909	AE013909 Yersinia
C 31	16.4	78.1	10530	1	AE012886	AE012886 Chlorobiu
C 32	16.4	78.1	198484	2	AC129435	AC129435 Rattus no
C 33	16.4	78.1	210050	1	AJ414146	AJ414146 Yersinia
C 34	16.4	78.1	22627	2	AC121370	AC121370 Rattus no
C 35	16.4	78.1	225929	2	AC109260	AC109260 Mus muscu
C 36	16.2	77.1	246	9	H006641S09	U36257 Homo sapien
C 37	16.2	77.1	366	1	AY366546	AY366546 Unculture
C 38	16.2	77.1	366	1	AY366549	AY366549 Unculture
C 39	16.2	77.1	366	1	AY366550	AY366550 Unculture
C 40	16.2	77.1	638	8	RFU30361	U30361 Rhodymenia
C 41	16.2	77.1	639	8	RFU30359	U30359 Rhodymenia
C 42	16.2	77.1	639	8	RFU30360	U30360 Rhodymenia
C 43	16.2	77.1	659	8	RFU30362	U30362 Rhodymenia
C 44	16.2	77.1	936	8	CRZY81A	X59683 C.reinhardt
C 45	16.2	77.1	1160	8	CRZY81B	X76117 C.reinhardt

# ALIGNMENTS

RESULT 1.  
AE006945/c  
LOCUS AE006945  
DEFINITION Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001  
complete genome.  
ACCESSION AE006945  
VERSION AE006945.1  
KEYWORDS GI:13879900  
SOURCE Mycobacterium tuberculosis CDC1551  
ORGANISM Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
REFERENCE 1 (bases 1 to 15178)  
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Unayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 15178)  
Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Unayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 15178  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
198..1457  
/gene="MT0401"  
198..1457  
/gene="MT0401"  
/notes="similar to GB:L20897 SP:P33221 PID:304887 GB:U00096 PID:1736489; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoribosylglycinamide formyltransferase 2"  
/protein\_id="AAK44623.1"  
/db\_xref="GI:13879901"  
/translation="MIDGTEEQHEPTVRHERPAAPQDVRVRWMLGSAEPSRELAIALQGAIEVADGVGAPARIADOSVVVMTDABELTAVIRELPDLVTVTAASVDALDAEQADCELTLPNARAVCTADREGLRLAADQLGPTAPFWGSLGELQAVAHAGPELVVAGVAGSSVAGVNEPAGVAGQAGHVPOTGVGSPRCAESVVEFEVATLVVCSQNGPLGIFEFCAPIGRDADAGELESWQPKLSTAAALDAKSIARIVKALGREGVGLMINDGVADVTVCPSAKVTVSQRSLVFEOLARILGLAVDTLMSFGAARVINDHTAGAAVGAAPADALTGALGVPESDVITFRGLGLVALATAPEVAIAERAREVNASRLNVDSRE"  
1873..3093  
/gene="MT0402"  
1873..3093  
/gene="MT0402"  
/note="similar to SP:P55218; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="O-succinylhomoserine sulphydrylase"  
/protein\_id="AAK44624.1"  
/db\_xref="GI:13879902"  
/translation="WDESSTVPKALPDGVQATVGVGMLRGFBETAAMVILTSGYGAARAKSFAEGLDHVYSRYGNEITVPWQVTLIECAPAFATAGAAVFTSGALLAGDRILVARSFGVFCVVCSEILPFWQVTVFDGDDLQWERSALSVPTQAVFETPSNQSLDIAAVTELAAAGAKVLDNVFPLPQQGFPLGVDDVTVVSGTKHIDGQVRVLGGAILGDREVIDGPKLMRHTGCPAMSAFNWLLKGLLETLAIRVQHSNAAQRIABFLNGHPVSRVRYPLSPHQPDLAKRQMSGGTVTVTALDCPEDVAKQRAFEVLDKMLDIDISNNLDAKSLVTHPATTHTRANGPESGAAIGLGDGVVRISVGLEDTDDLADIDRAL"  
complement (3090..4502)  
/gene="MT0403"  
complement (3090..4502)  
/gene="MT0403"  
/note="similar to GP:2708705; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="NADH dehydrogenase"  
/protein\_id="AAK44625.1"  
/db\_xref="GI:13879903"  
/translation="MTLSGEPSSAVGRRHVVIGSGFGLNAKALKRADVDITLIS

XTTHLQPLLYQVATGILSEGDIAPTRILRLRQKNRVVLGEVNAIDLKQVTVSK  
LMDMTVTYDLSLVAAGAQSYFNGDEFATFAGMKTIDDAELRGRILGAFPAEAV  
STDAERERLTVVVGAGPTGVEVAGIIVELAEITLAGARITPTTSECRIILDAAP  
AVLPMGPGLKQORLEKMDVEVLNAMVAVDYKGITIKEDGGERIECAKCMW  
AAGVAAPLGMIAEGSDGTEDRAGRVIVEPDLTVKGHNVFVVGDMFVPGVGA  
QGAIOGARYATTVTKMVKGNDDPANRKPHEYNKSGMATISSHSAVOVKLEFAGY  
FANLAWLVHLVLYVYGRNRIALFANGISFMGRARGMAITSQMITYARLVNTLMEQO  
AQGALAAABAEQAHAQSAAG"  
4644..5969  
/gene="MT0404"  
4644..5969  
/gene="MT0404"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44626.1"  
/db\_xref="GI:13879904"  
/translation="MAVGRCAIPFDOAASGSAINGGVHLSDGSTSPARQLPAPWPG  
DAGAAEGRAGVCCGRNLRPHVSDVGVSHRFDHRPAGVAGGCGRAGAGLAVDDPG  
CLAAADRIVAVADPDVAVRQVRERARDREVISMNSADMGGEVYLAQLTATDAALDAKL  
NALVATVCGDPRSTDQRDALGALAAAGADRLACRCDNPDCAEGRPSAVVHVVA  
EQASVKGHGQAPALGLGDLIPAEVLAELAKTAGLQFIIPVPAETGEPGSPVLAFA  
VRADLTCTRAPGCDRATQCDLHTIAFADGGATHAANLKLRLHLLATFCGWRAQ  
QLPDGTVITLPGNQTVTTTSGSALLFPALCTPTGDDPAPEPARADRRGORTAMPRR  
ASTETQNRACIAERHRHQARRIAQAIAVIAETHGPPDPDDDDPPPF"  
complement (5985..6704)  
/gene="MT0404.1"  
complement (5985..6704)  
/gene="MT0404.1"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44627.1"  
/db\_xref="GI:13879905"  
/translation="MTEPRFVAVVIAIGLSAIFPMVGPLQTVDFDAIEERTRAEFTT  
TRECSEVGADTVLRIDKNPELEPLLSQIAEATTSMEAKRRLIAQAAALAEED  
TRECSEVPSLIVATLSQLEPRLHVLRLAKAASQSPODEIORREVMAASKVPVPL  
AALITQGVATATTVHNGTGPAAESSGHILIHVDVDFGHRLLAVLRADAGALLI  
LPSCGSAPTGDHPTPHFSTSR"  
6803..7207  
/gene="MT0405"  
6803..7207  
/gene="MT0405"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44628.1"  
/db\_xref="GI:13879906"  
/translation="MDMVLPGDYETFRHWSGKPRAMPQBSGWRWFVGGKIVDGLCEV  
LDEHLAVRRGVPAAIIGCVFWLSAEVAETLLALSAPCVVIDKGTSPFSLRNPDKGF  
PNVALLRLRMAPSEHSGSRCCSSARGRLCLMS"  
7189..7605  
/gene="MT0406"  
7189..7605  
/gene="MT0406"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44629.1"  
/db\_xref="GI:13879907"  
/translation="MPEHELGPVRLGWLREDRKPLINAKLLVLGLHNLNVDPNGY  
GEEVLPERTVWMSANTVRAGSHLEVGFACDDPTLVEEATAFVADVIXXSEIXDT  
TCAGPEXNLVQVRFDDAAMAEAMEAEPPDDGCDW"  
complement (7646..8116)  
/gene="MT0407"  
complement (7646..8116)  
/gene="MT0407"  
/note="identified by Glimmer2; putative"  
/codon\_start=1



```
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4630.1"
/db_xref="GI:13879908"
/translation="MKYNNIRTPCLYPIPVFWAGAYRLRGRVYVVGIGWAVSLGGGCG
GLGDHVGNPAGLVAVAFGQVVGAVLRSGAGPARHHRGALPAAVAGRGVSGRWVWG
GGAORRQORAAAGGDIQLVTRQPNHGAUGQLGAPAPGEGGQWMAPIITRTF"
8257..8505
/gene="MT0407.1"
8257..8505
/gene="MT0407.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4631.1"
/db_xref="GI:13879909"
/translation="MHALLVGLAILTAIPIAVLIGSSPAHADTDIGQCPSPGAKL
WNGPGIYCERTADGQLOWVSIPIAWALCVAFCDRGGP"
complement(8542..9183)
/gene="MT0408"
complement(8542..9183)
/gene="MT0408"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4632.1"
/db_xref="GI:13879910"
/translation="MGVIARVGVGAAGLSLAVLAAPTAGAEPFGALPPMTSSGGP
VIGDPAALQRISQOLFSGDPTQVQVDSGAQPTTAAAVADRDVASVFLPIQV
LGCQNTAGSGAGFARAYRTDQWGAMLVVAKTSDVDALACVKGWRKATAG
TPTSMCNGWTVPFPADTRGEGYFVLLAGTASDFCSAPNANYRTTASSWPG"
complement(9190..10419)
/gene="MT0409"
complement(9190..10419)
/gene="MT0409"
/notes="identified by match to Prosite PS00013"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
100.0%; Score 21; DB 1; Length 15178;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTGGCAGCGCCCAACAGC 21
DB 4667 GATTGGCAGCGCCCAACAGC 4647

RESULT 2
BX248335/c
LOCUS
DEFINITION
Myxobacterium bovis subsp. bovis AF2122/97 complete genome; segment
2/14
ACCESSION
BX248335 BX248333
VERSION
BX248335.1 GI:31617046
KEYWORDS
complete genome.
SOURCE
Myxobacterium bovis subsp. bovis AF2122/97
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
REFERENCE
AUTHORS
Garnier,T., Eiglmeyer,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Morsempe,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
TITLE
The complete genome sequence of Mycobacterium bovis
JOURNAL
Online Publication
REMARK
FNAS 10.1073/pnas.1130426100 ( Microbiology )

REFERENCE
AUTHORS
Garnier,T.
TITLE
Direct Submission
JOURNAL
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton,Cambridge CB10 1SA, UK. F14 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
FEATURES
Location/Qualifiers
1..324050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
132..2027
/locus_tag="Mb0290"
132..2027
/locus_tag="Mb0290"
/notes="Mb0290", len: 631 aa. Equivalent to Rv0282, len:
631 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.8% identity in 631 aa overlap). Conserved hypothetical
protein, similar to Y14967|MLCB628.18c hypothetical
protein from Mycobacterium leprae (573 aa), FASTA scores:
opt: 916, E(): 0 (38.7% identity in 568 aa overlap). Also
similar to Mycobacterium tuberculosis proteins e.g.
294121|MTY15F10.26 (619 aa), FASTA scores: opt: 743, E():
0 (29.9% identity in 612 aa overlap). Member of CFXP,
CBXP family - 9 members in Mycobacterium tuberculosis.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD93154.1"
/db_xref="GI:31617047"
/translation="MAGVAGDSGGVERDDIGMVAASPVASRVNGKVDADVGRFATC
CRAIGLVYQKRPDPAALAAARSFAALTTRVAHDOCDADWTGLAAAGDSIGGVLEAASRT
ATTAGVLQROVELADNALGLYDGLVLRFRATGPDPPHLYAALASTGSGPEEPFAPA
NHVSGTERRAGRAARNLAVINRAERWSVDVVKLLTPMNPDPDLDEASHAKIT
LGTALRGMPALSYLEEDPGVAAVVDGAKALVLRHVDDEASASVLOQLYA
APNEQVEALSDFIVTTTARTIEARTDPDPEAGDFVDPFAAHERKAALL
HEALQIAEFTGLDEKQVSRKSSVAMELVKQKGLTVAQTHLVFAGPGTGKT
TIARVAKIYCGLLKKREIRVHRADLIGHIGETEAKTNAIDSDGLFIDEA
VALVATGAKNDFGLVADTLIARMENDRDLVLIAGYRADLKDPLDNEGLRGRFTR
NIDPSTYTSHELVEIAHMAQRDSVPEQSALHDLALFALAEPTDITGSRSL
DIAGNRFVRNIVERSEEREFRLDHSHAGSEFSEELNTIITADVGRSVPELIRG
LGLSVRA"
2024..3640
/locus_tag="Mb0291"
2024..3640
/locus_tag="Mb0291"
/notes="Mb0291", len: 538 aa. Equivalent to Rv0283, len:
538 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 538 aa overlap). Possible conserved
membrane protein, similar to several hypothetical
mycobacterial proteins e.g. Z94121|MTY15F10.16|Rv3895c
from Mycobacterium tuberculosis (495 aa), FASTA scores:
opt: 698, E(): 0 (37.6% identity in 492 aa overlap);
Rv1782; Rv3450c; Rv3869; and Y14967|MLCB628.16|MLCB628.17c
from Mycobacterium leprae (481 aa), FASTA scores: opt:
672, E(): 1.5e-31, (37.2% identity in 506 aa overlap).
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD93155.1"
/db_xref="GI:31617048"
```

/translation="MTNQOHDHDFDHRRSFASRTFVNNPNPKVYVRGFTVRHQVTG  
WRFVRIAGIALHDFRMLVDRLTQSRVAVLGLVITLGLIGSVFSLIRPNGQAG  
SNAVLADRSATLAVRVEGOLHVLNLTARLTVGRPVSTTVKSTELDFQFRNLIG  
IPGAPERVQNTSTNDANTWCDGLNAPSRGAGDGVGVYIAGLEDTGARAALPGQ  
AVLVDSGAGLLWDGKRSPLDIADRAVTSGLGLGADVPAPRIIAGSLNAPAPPL  
TAPIIDPAGNPFAGVPAPIGAVSSYALKDSKTIISDTQVYAVLPDGLQIISPVLA  
AILLNNSYGLQPPRIGLGADEAVKLPSVRLDTRRPSEFVSLVDRDFTVCAYWSK  
PVGATSSLLTAGSLPVPDAVHTVELVGAGNGVATRVALLAAGTGYFTQTVGGGPD  
APGAGSLFWSDTGVYIGINEQGVAGGKGAVALGLNPPVPPIFMSVLSLFPVPGPT  
LSRADALLADTLVDSRPARPVSAEGYR"

gene

3637.7629  
/locus\_tag="Mb0292"

CDS

3637.7629  
/locus\_tag="Mb0292"

/note="Mb0292", len: 1330 aa. Equivalent to Rv0284, len: 1330 aa, from Mycobacterium tuberculosis strain H37Rv, (99.9% identity in 1330 aa overlap). Possible conserved membrane protein, similar to products of two adjacent Mycobacterium leprae genes, MLCB628.16c (744 aa) and MLCB628.15c (597 aa); and throughout its length to several large Mycobacterium tuberculosis proteins: Rv3447c, Rv3870, Rv1784, etc. Y14967|MLCB628\_15 (744 aa), FASTA scores: opt: 942, E(): 0, (33.8% identity in 730 aa overlap); Y14967|MLCB628\_14 (597 aa), FASTA scores: opt: 613, E(): 3.1e-30, (31.7% identity in 615 aa overlap); Z94121|MTV15P10.17 (1396 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap); Z95389|MTCY77\_19 (1236 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap). Contains three P800017 ATP/GTP-binding site motif A (P-loop)."

/codon\_start=1  
/evidence=experimental

/translation="MAAPINWASPEVHSHALLSNGPGPSILVAAATWSQLSAEYAST  
AAELSGLLGAVPGWAGQPSAEVVAHLPYAWLTQASADAAGAAAEHAAAAYTT  
ALAAAMPTLAEALAHVHTVLVATNFITITLNEADYVRMLQAAMVGLYQAA

/transl\_table=11  
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"

/protein\_id="CAD93156.1"

/db\_xref="GI:31617049"

/translation="MSRLIFEARRRRLAPSSHQGTIIIEAPPELPRVIPPISLLRRALP  
YLIGLIVGMVALVATGMRVSPQITLFPFFULLAANTALYRGNDKXKTEBDAERA  
DYRLYSVRNDNRQAQASALWSHSDPTLALSPGSRQWERDHPDPLFLVLA  
GERVPLATLRVNDTAEIDLEFVSHSALRSLDTSIGVFTGIDITKVSRIVL  
GERAQRVLRVIAQAVTWHDTPLVGLAARDEGRDNWMLKPLHPIVDFGLADL  
GPARNLSTDELIALLGPLVADRPAFTQPTDALRHLLTVDDPDYDLGASPLAVGR  
AGTVVHCASAPHEQYSDPEKPIRLVAHGERWOTGCOPYIDAQDFSADEBAH  
LARRLSRNSNTHAGLSAATRGASFTTLGIEDASRLDVPALNAPRRERDELVPI  
CVTGTGEPNLFDKDAEGMGPHGLMIGTSGKSKQTLMSILLSTHSERLVI  
YADFGBAGADSFRDPFQVAVVLSNMAERKSLADRFADLRGEVAREMLLRAGKV  
QGSFNSVLEYNAIAGHSLPITPLFVVADEFTLMDADHPEALFHYVARKGRSF  
RIHLFASQTLVDGKIDKIDNTAYRIGLKVASPSVSRQIIGVEDAYFYSKEHGV  
GFLVPAGATPIRSTYVDGIVEPQTAKAVVQSVPEPKLPTAAAVPEPDTVIAD  
TDQEPADPRKLIATIGOLARYGPAPOLPLDDETIPLSALARAGVGPROMRV  
PLGEIDRPFMRDPLVDFARSAGNMVHGGKSGKSTALOTFILSARLSHPEVS  
FYCLDGGGQLADLHAGVSVASALEPERIRRTFGELEQLLSRQREVERDRGAN  
GSTPDGFGVFLVONLYGFRDNDQFNTRNPLIARVDELVNGLAYGIHVIITTP  
SWELEPLAMDGLGLELRDLHSDNSNVVGLARPADAPHDQPGRLTMAAHPF  
LFAPELADQTNVPAINARYPGMAAPPVLLFTNLPAHAGVGLRGPDLVQIGORE  
DLAPVILDAAANPLAVFGDAGSKTLLRHITVRHSTDRVATVLDRLRLHVO  
EPLFPNEYNTADIRIIPAMGLANLIEARRPAGMSAAELSWTFAGHTHYLIDDV  
DOVDPSPMTGPVIGORPWTPLIGLAQDGLRVLVTGRATGSAHLMTSPLARRF  
NDLQATMLMAGNPADSGKIRGERFARLPAGRALLLTSDSDSTFYQLINPLVDAAVS  
GETQQKGSQ8  
7626.7934  
/gene="PES"  
/locus\_tag="Mb0293"  
7626.7934  
/gene="PES"  
/locus\_tag="Mb0293"  
len: 102 aa. Equivalent to Rv0285, len: 102 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 102 aa overlap). Member of the Mycobacterium tuberculosis PE family (see first citation below), similar to others e.g. AL0212|MTV012.37 from Mycobacterium tuberculosis (105 aa), FASTA scores: opt: 497, E(): 2.6e-24, (80.4% identity in 102 aa overlap);

gene

7626.7934  
/gene="PES"

CDS

7626.7934  
/locus\_tag="Mb0293"

7626.7934  
/gene="PES"

7626.7934  
/locus\_tag="Mb0293"

/note="Mb0293", len: 102 aa. Equivalent to Rv0285, len: 102 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 102 aa overlap). Member of the Mycobacterium tuberculosis PE family (see first citation below), similar to others e.g. AL0212|MTV012.37 from Mycobacterium tuberculosis (105 aa), FASTA scores: opt: 497, E(): 2.6e-24, (80.4% identity in 102 aa overlap);

Z80108|MTCY21B4.03 from Mycobacterium tuberculosis (102 aa), FASTA scores: opt: 413, E(): 3.7e-19, (66.7% identity in 102 aa overlap); etc."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="PE FAMILY PROTEIN"  
/protein\_id="CAD93157.1"  
/db\_xref="GI:31617050"  
/translation="WTLRVVPEGLAASAAVEALTARLAHAASAAVITAVVPPAAD  
PVSLQTAAGFSAQGVHVAVTAGVEELGRAGVCGESGASYLAGDAAAATYGVVGG  
"

gene

7937.79478  
/gene="PPE4"

CDS

7937.79478  
/locus\_tag="Mb0294"

7937.79478  
/gene="PPE4"

/locus\_tag="Mb0294"  
/note="Mb0294", PPE4, len: 513 aa. Equivalent to Rv0286, len: 513 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 513 aa overlap). Member of the Mycobacterium tuberculosis PPE family, similar to others e.g. AL0212|MTV012.32 from Mycobacterium tuberculosis (434 aa), FASTA scores: opt: 958, E(): 0, (43.5% identity in 522 aa overlap)."

/codon\_start=1  
/evidence=experimental

/transl\_table=11  
/product="PPE FAMILY PROTEIN"

/protein\_id="CAD93158.1"

/db\_xref="GI:31617051"

/translation="MAAPINWASPEVHSHALLSNGPGPSILVAAATWSQLSAEYAST  
AAELSGLLGAVPGWAGQPSAEVVAHLPYAWLTQASADAAGAAAEHAAAAYTT  
ALAAAMPTLAEALAHVHTVLVATNFITITLNEADYVRMLQAAMVGLYQAA

Query Match 100.0%; Score 21; DB 1; Length 324050;  
Best Local Similarity 100.0%; Pred. No. 12;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGCGCAGCGCCCAACAGC 21  
|||||

Db 130823 GATTGCGCAGCGCCCAACAGC 130803  
|||||

RESULT 3

BX842573/c

LOCUS BX842573 342416 bp DNA circular BCT 21-NOV-2003  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 2/13.

ACCESSION BX842573 AL021930 AL021931 AL021932 AL021933 AL021942 AL123456

277162 Z83865 Z84724 Z84725 Z95324 Z95558 Z96796 Z96800 Z97831

297991

VERSION BX842573.1 GI:38490165

KEYWORDS complete genome.

SOURCE Mycobacterium tuberculosis H37Rv

ORGANISM Mycobacterium tuberculosis H37Rv

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

REFERENCE 1

AUTHORS

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,  
Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,  
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K.K., Skelton, S.,  
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., and  
Barrell, B.G.

TITLE Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence

JOURNAL Nature 393 (6685), 537-544 (1998)

MEDLINE 98295987

PUBMED 9634230

```

REFERENCE 2 Camus,J.C., Pryor,M.J., Medigue,C. and Cole,S.T.
AUTHORS Re-annotation of the genome sequence of Mycobacterium tuberculosis
TITLE H37Rv
JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
MEDLINE 22255591
PUBMED 12368430
REFERENCE 3 (bases 1 to 342416)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT On or before Nov 21, 2003 this sequence version replaced
gi:3261524, gi:3261526, gi:3261527, gi:3261529, gi:3242298,
gi:3261606, gi:3261689, gi:3261708, gi:3261703, gi:3261760,
gi:3261781, gi:3261797, gi:3261800, gi:3261825, gi:3261837.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).
FEATURES
Source
1..342416
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/db_xref="taxon:83332"
223..2118
/locus_tag="Rv0282"
223..2118
/locus_tag="Rv0282"
/locus_tag="Rv0282"
/functions="UNKNOWN"
misc_feature
/translation="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAA17358.1"
/db_xref="GI:2909472"
/translation="MTNQOHDHDEHRRSFASRTPVNNNDKVVYRGFVTRHVTG
WRFVNRRAAGIALHDTLPRTQSRVLMGVLIITGLIGSFSLRPNQOAG
IPGAVRMTQNTSDYRVGEQLHPVLTARLIVGRPVSTPTVSTELDQPPKGLG
AVLVDGAGTWLLMDGKSPIDLDADHVTSGLGADVPAPRIITAGLFLNAPEAPPL
TAPILPDAGNPASFGVPAPIGAVYSSYALKDSGKTIISTVQYVAVLPGLQILSPVA
ALLRNNISGICQPRGADVAKLVSRVLDTRKIPSEPVLDVDRDPVTCATWSK
PVGAGTSLLTLAGSALPVPDAVTHVELVAGNGVATRVAAAGTGYTQTVGGGPD
APGAGSLFVSDTGYRIDNEPQVAGGKAVKALGNLPPVPPIPSVLSLVEGPT
LSRADALLAHDTLVDSRPARPVSAEGGR"
3024..3047
/locus_tag="Rv0283"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
3728..7720
/locus_tag="Rv0284"
3728..7720
/locus_tag="Rv0284"
/locus_tag="Rv0284"
/functions="UNKNOWN"
/notes="Rv0284, (MTV035.12), len: 1330 aa. Possible
conserved membrane protein, similar to products of two
adjacent Mycobacterium leprae genes, MLCB628.16c (744 aa)
and MLCB628.15c (597 aa); and throughout its length to
several large Mycobacterium tuberculosis proteins:
Rv3447c, Rv3870, Rv1784, etc. Y14967|MLCB628.15 (744 aa),
FASTA scores: opt: 942, E(): 0. (33.8% identity in 730 aa
overlap); Y14967|MLCB628.14 (597 aa), FASTA scores: opt:
613, E(): 3.1e-30, (31.7% identity in 615 aa overlap);
294121|MTY15F10.17 (1396 aa), FASTA scores: opt: 652, E():
2.2e-32, (35.4% identity in 1321 aa overlap);
295389|MTCY77.19 (1236 aa), FASTA scores: opt: 652, E():
2.2e-32, (35.4% identity in 1321 aa overlap). Contains
three PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAA17359.1"
/db_xref="GI:2909473"
/translation="WSRLIFEARRLAPPSSHQGTIIIEAPPELPRVLPSSLRRALP
YLIGLIVGMIVALTGMRVISQTLFFPVLIIAATLYRGNDKXMRTEBVDARA
DYRLYSVVRNIRQAQAEQASALWSHFDPTALASVPSRRQWERDHPDFVIVRA
GRHTVLTATLRVNDTADIEDLPVSHSALSLDQTSRIGDPTGIDLTQVPTVL
GERAVRLVAIAQVATWHTDPTVIGVALAARDLEGRDWNKWLPHVDIIPGRDLAL
GPARNLSDPBELIALGLVADRPATFQPTDLRHLLIIVDDPDYDLGASPLAVGR
AGRTVVCASAPAHREOYSDPEKPLIRVAHGAIERWGTQWQVPIIDAADQFSADBAH
LARRLRNDSNPTHTAGLSAATRGASFITLIGIEDASRLDYALVAPRDEELRVI
GVTGCEPMTDKDEAGSGPGLMIGTSGSKQSLMISLLSLTHSAERLVI
YADFGEAGASFRDFPQVAVIINMAEKKSLADRFADTLRGEVAERLELLEAGKRV
QGSASFNVLEYENAIAGHSFPITPLEVFADEFTMLADHPEVAELFDYVAKGRSF
RIHLFASQTLQDVGKIDIKDNTAYRIGLVKASPSVSRQIIGVEDAVHIESGKEHGV
GLFVAPGATPIRFRSTVDGIIYEPPTAKAVVQSVPEKPLTAAVPEPPTVIAD
TDEQSPADPRKLIATICEQIARVPAPQLWLPDLDETIPLSALAGVGPROMRW
FYCLDGRPFEMRRDPLVDARSAGNVIHGGKSGKSTALQTLFSASLSHSHEVS
GSTPDGFGVFLVDINLYGFRDNTQFNTRNPLLARVLELVNVLNGLAYGHVITTP
SWLEVPFAMRGDLGLRLRLRDLNDQNVVVGALRPADVPADPQGRGTMAAHIFP
LFAPELDAQNPVAAINARYPGMAAPFVRLIPTNLAPHAVGELYRQDVLVIGORE
DLAPVLDLAANPLMVFGDARSGKTLRLHIRTREHSTADRVAFVLDRLRLHVD
EPLPDPNEYTANIDRIIPAMIGLANLIEARPPAGMSAELSRWTFAGHTHYLIIDVV
DOVPSPMATGPYICQRWTPILIGLAGAGDLGLVITVTRGATGSAHLLMTSPLLRRF

```







TITLE Direct Submission  
 JOURNAL Submitted (13-DEC-2002) Department of Microbiology, Genome Research Center for Enteropathogenic Bacteria, Chonnam National University Medical School, Hak-I-Dong, Dong-Gu, Kwang-Ju 501-746, South Korea

FEATURES  
 Location/Qualifiers  
 1. .301235  
 /organism="Vibrio vulnificus CMCP6"  
 /mol\_type="genomic DNA"  
 /strain="CMCP6"  
 /db\_xref="taxon:216895"  
 /chromosome="I"  
 182. .532  
 /locus\_tag="VV10587"  
 182. .532  
 /locus\_tag="VV10587"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Conserved hypothetical protein"  
 /protein\_id="AAO09103.1"  
 /db\_xref="GI:27360166"  
 /translation="MEARYVLLRVETLRTVASRPVRSYRTLSPLPDPPLIIRIKAIAG  
 GLLSVALVGLRPPGVTHPALWSPDFPLCQSPGRQSPKGLQRQSPGEGHQSSD  
 YFVNSGGLYRLIH"  
 complement (602. .1465)  
 /locus\_tag="VV10588"  
 complement (602. .1465)  
 /locus\_tag="VV10588"  
 /notes="COG0313"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Predicted methyltransferase"  
 /protein\_id="AAO09104.1"  
 /db\_xref="GI:27360167"  
 /translation="WTDKSNFTEGATLIIVPTPIGNLGDITQRAIEVLSVDMIAAE  
 DTHGKLHFNITKLTALHDNEQQAQVILVDKLLSQNIALVSDAGTFLIDSPG  
 YHLVTCQAGVPLPGACAVITALSGLPSDFSGFGLPPKSGRKRKDFLEIA  
 KVERTICFYESPRIVDSLHMLEILGPRVYVLAELTKFTETIQMELSLVWVQ  
 SDNOQKGEMVLLIHGRDSABESLPEALRSALITKELPLKKAALAAETYNLKN  
 ALYKGLNLE"  
 1534. .3345  
 /locus\_tag="VV10589"  
 1534. .3345  
 /locus\_tag="VV10589"  
 /notes="COG3107"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Putative lipoprotein"  
 /protein\_id="AAO09105.1"  
 /db\_xref="GI:27360168"  
 /translation="MNPKELSVPRLLTPVALAITLAACSGPKTPTSVDITILEPTLS  
 VQNYMINADSSGSLQADMLIMAAKALQNGDLAQDLILIKRLMPLSEVQAEWQL  
 TRAYHLKNOADNVIELNFKAWMLPNEQWQYVTLRAEAYTALQOQPEANRLVLT  
 LSOYVDESQOAEIAQIWNFTCYSOYEITOLHPDASEVLDGWLQALATYMTMSANY  
 PQKNTLEHFTENTHPRALVTPARIQSLTLDLEIVQVHTALLPLSGYAKQROLI  
 RDGFIFAMNDKDRDEATLKVIDTNLYQPHLQKQLTBQIDDFIVGPRKQVIEVLQ  
 GELSDSGQVYPSLAINPDELQGTGTCYLTLSPEQVAQAHLFANGYKYPILF  
 APQGNLCQVRSFAEAWKRYSTNVAVSYFGKQROLQDVNSVFGLSQSQRIQME  
 GLMKLPMTQVRRRDIQSVIYAARSSELTLIKPFIEVAVNPDAKPKLFSNMSNSG  
 EKYEDLTGTVSDIPMLLEVPALDSQMEQLWPDQSNFQKRLQALGMDAYKLMAELP  
 QMKVVPNHNANGTGVLTIDQCQVHREISWKEGAL"  
 3332. .3700  
 /locus\_tag="VV10590"  
 3332. .3700  
 /locus\_tag="VV10590"  
 /notes="similar to to archaeal Holliday junction resolvase;  
 COG0792"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Predicted endonuclease"  
 /protein\_id="AAO09106.1"  
 /db\_xref="GI:27360169"  
 /translation="MGLFSRRAIGNQVESLAKEYLQROGLRFTFEANFTTKVGIDILIF  
 KEAQTIIVFEVKYKNSCYGDAAEWNPAPKANLKITAYLWLNKHGYNACNTAMRPDY

VAHSNGHDINWIANAITQG"  
 3704. .4294  
 /locus\_tag="VV10591"  
 3704. .4294  
 /locus\_tag="VV10591"  
 /note="COG0279"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Phosphoheptose isomerase"  
 /protein\_id="AAO09107.1"  
 /db\_xref="GI:27360170"  
 /translation="MLDSIKDSFTESIQIAAAALPDAITHAAQAQAVATLLNGHKI  
 LCCNGGSAANVOQFVSLNRPETEPESLPAMALTADNTTLTAVANDYHQEIFSKQ  
 VRAFQPGDILLAIESTGNSKNIIKAMEAAVTRDMTIIALTGDKGEMAGLIGENDVE  
 IRIPSHRTARIHEVHMVTLHCLDLIDQVLPFAHEE"  
 4297. .4890  
 /locus\_tag="VV10592"  
 4297. .4890  
 /locus\_tag="VV10592"  
 /note="COG2823"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Predicted periplasmic or secreted lipoprotein"  
 /protein\_id="AAO09108.1"  
 /db\_xref="GI:27360171"  
 /translation="MLRYVLLFITWALSGCAGLFIAGAATANLVTDTRTKEIWQD  
 NNIFEJAAIGNKAPYRGKVIIVASSYNGTVLWLGQASTSEPSGFEQOASIKGVQV  
 LHNQVKAPLSVSOISNDSWITTKVKSALLTDESLSVKIKVITEDGDVFLFGVNS  
 EQAEKAEIARNISGVQVVRAPQYGENATQVNNASQ"  
 complement (4967. .5443)  
 /locus\_tag="VV10593"  
 complement (4967. .5443)  
 /locus\_tag="VV10593"  
 /note="COG2969"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Stringent starvation protein B"  
 /protein\_id="AAO09109.1"  
 /db\_xref="GI:27360172"  
 /translation="WEIAQMTARRPYMLAFYDWLVDNDLTPHLVVOATLAGVRVQPE  
 FIQDQIITINIPRAVGNLEIGNDTISFNARFGRGPHSVIPLVAVLAIYARENGAGT  
 MFPEEATYSPGAEQEBDDGIIWLEEHSAQPSIDFDEPEPPKRPKPSLRVVK"  
 complement (5447. .6082)  
 /locus\_tag="VV10594"  
 complement (5447. .6082)  
 /locus\_tag="VV10594"  
 /note="COG2969"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Stringent starvation protein A"  
 /protein\_id="AAO09110.1"  
 /db\_xref="GI:27360173"  
 /translation="MALAANKRSVMTLFSSASDLYSHQVRIVLAEGSVSEVELVDET  
 NLPLIELNPKVSPVTLVDRELALYDSKIIMEYLDERFPPHPLMPVVPVARGNSRLM  
 MYRLERNVYSLAEKIVKGNABESKARNKLANDLLTSPFAEFYFMSEFSLIDCY  
 LAPLLWRPLGIDILIGPSKEIKAYMKNKVERFSFLASLEAREMELVR"  
 complement (6180. .6917)  
 /locus\_tag="VV10595"  
 complement (6180. .6917)  
 /locus\_tag="VV10595"  
 /note="COG2857"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Ubiquinol-cytochrome c reductase, cytochrome c1"  
 /protein\_id="AAO09111.1"  
 /db\_xref="GI:27360174"  
 /translation="MKKWIVILFAMLPISLMAAGAGVPLDKANVLDSDKASIQNGAKI  
 FMNYCFACHSTQYQRYERADDLGI PADLMKENLMFNPSKIGDLNVNMPAKOAAAW  
 FGAPPDLTILVARVGADWLTYLTFTFYVDSRSPGVNNLVPPSPVGVNPHVLEELQIP  
 NPVYTKVVDGREVQVVGVSVDGTGELSGKEIKAVLDLVSFLVYSGDPVKAERHAL  
 GWNWAFVLVLTIVVLLKKEYVRDVH"  
 complement (6914. .8179)  
 /locus\_tag="VV10596"







[illegible]

EHOXKLEQTEVEMKOKETAAKQOELTDONVDSLSEKLSSENLIHDVLOREKEKMEKE  
 NDYLXRENNRLLSYMGDLDSLSYQVFTSNQGTRODQPLGVFGNPSYGCQESSAPSL  
 SIKCKPTCKSDFAADVDHNLALAEQHLQTLSLNCPICDKTFFPAKEXQIFEDHVFTCL  
 "

polyA\_site  
 2196  
 /gene="NDP"  
 /note="21 A nucleotides"

## ORIGIN

Query Match 80.0%; Score 16.8; DB 4; Length 2196;  
 Best Local Similarity 90.0%; Pred. No. 1.8e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAG 20  
 |||||  
 Db 1879 GATTGGCAGCTCCCAACAG 1898

RESULT 11  
 AF059613  
 LOCUS 2809 bp mRNA linear INV 07-JUL-1998  
 DEFINITION Drosophila melanogaster transportin mRNA, complete cds.  
 ACCESSION AF059613  
 VERSION AF059613.1 GI:3293341  
 KEYWORDS  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 2809)  
 AUTHORS Sioni,M.C., Fromont,M., Rain,J.C., Wan,L., Wang,F., Legrain,P. and Dreyfuss,G.  
 TITLE Functional conservation of the transportin nuclear import pathway in divergent organisms

JOURNAL Mol. Cell. Biol. 18 (7), 4141-4148 (1998)  
 MEDLINE 98298258  
 PUBMED 9632798  
 REFERENCE 2 (bases 1 to 2809)  
 AUTHORS Sioni,M.C., Fromont,M., Rain,J.-C., Wan,L., Wang,F., Legrain,P. and Dreyfuss,G.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-APR-1998) Biochemistry and Biophysics, HMI and University of Pennsylvania School of Medicine, 415 Curie Blvd., Philadelphia, PA 19104, USA

FEATURES  
 source Location/Qualifiers

1. 2809  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 24. 2705  
 /note="TRN"  
 /codon\_start=1  
 /product="transportin"  
 /protein\_id="AAC25708.1"  
 /db\_xref="GI:3293342"  
 /translation="MTWEPQEGEQIIAILKESQSDPTATQNAVMKLEEFNRYPDF  
 NNLYIYVLTKEDEFTSLGSLIKNNIRHGTTLQPEIYEIKHECLQAVGSSP  
 LIRATVGLITTIASNGLNHPQLPSLCLMDNQDYNVCEGAFSALQKICEDSAEI  
 LDSAALNPLNIPKFLYFKHSPKIRSHAIACINQFIINRSQALMLNIDSFIEL  
 FHLSSDEHVRNGLVGLVLLKVRMDLPHMSQIIEYMLLRTDQDGVLEBASE  
 FWLSLAQCTKDVLAAPYLAQAPLVARGMYSEVDIILLKGNVEDDMVPREDIR  
 PREKSRATHIRSTQEGGAGATGDDDDFEFGMDSSLSSENLKCSAAALDVLAN  
 VFREDCLPIVPLIKETLPHQEWIKESGVALGALAEQCMQOHLPFLIPYLISC  
 LSDKALVRISITLUSRVANVWVQPHDQYILKPLMEKLLKILDSNRKRVQEAACSF  
 ATLEEACETLVPYLYIKTLVFAFSKYQHNKLLLYDAVGLSDSVHGLNKRPQYI  
 DIIMLEPIIDKDDKDLFPLECLSLIATFALQSGFLPCDPYRRCILIRQTI  
 NQMLCKQNTYDHPKERMIVALDLSLGAELDRHIELTVLGNNSIMHLLYQCMQDV  
 LPVROSSFALLGLTKACFPVHPFMADFFPIILGONLNPDISVCNNATWALGICM  
 KLGEEFKOYRLVLSOLFILNRPNTPKLLNTATIGELGVCEVAVPYLPEFVR  
 QWCTSLRHINDNDEKSAFRGCMCHMTIVNPAGVADFIFPCDAIASWNPFDRLRQMI  
 QKLLHGFKTKQVGENWCRFVEQFPPTLAERLITWYNI"

## CDS

CDS  
 1. 3377  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="Y; cn bw sp"  
 /db\_xref="taxon:7227"  
 1. 3377  
 /gene="Trn"  
 /db\_xref="FLYBASE:FBgn0024921"  
 147..2828  
 /gene="Trn"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RES9670p"  
 /protein\_id="AAO24946.1"  
 /db\_xref="GI:27819795"  
 /db\_xref="FLYBASE:FBgn0024921"  
 /translation="MTWEPQEGEQIIAILKESQSDPTATQNAVMKLEEFNRYPDF  
 NNLYIYVLTKEDEFTSLGSLIKNNIRHGTTLQPEIYEIKHECLQAVGSSP  
 LIRATVGLITTIASNGLNHPQLPSLCLMDNQDYNVCEGAFSALQKICEDSAEI  
 LDSAALNPLNIPKFLYFKHSPKIRSHAIACINQFIINRSQALMLNIDSFIEL  
 FHLSSDEHVRNGLVGLVLLKVRMDLPHMSQIIEYMLLRTDQDGVLEBASE  
 FWLSLAQCTKDVLAAPYLAQAPLVARGMYSEVDIILLKGNVEDDMVPREDIR  
 PREKSRATHIRSTQEGGAGATGDDDDFEFGMDSSLSSENLKCSAAALDVLAN  
 VFREDCLPIVPLIKETLPHQEWIKESGVALGALAEQCMQOHLPFLIPYLISC  
 LSDKALVRISITLUSRVANVWVQPHDQYILKPLMEKLLKILDSNRKRVQEAACSF  
 ATLEEACETLVPYLYIKTLVFAFSKYQHNKLLLYDAVGLSDSVHGLNKRPQYI  
 DIIMLEPIIDKDDKDLFPLECLSLIATFALQSGFLPCDPYRRCILIRQTI  
 NQMLCKQNTYDHPKERMIVALDLSLGAELDRHIELTVLGNNSIMHLLYQCMQDV  
 LPVROSSFALLGLTKACFPVHPFMADFFPIILGONLNPDISVCNNATWALGICM  
 KLGEEFKOYRLVLSOLFILNRPNTPKLLNTATIGELGVCEVAVPYLPEFVR  
 QWCTSLRHINDNDEKSAFRGCMCHMTIVNPAGVADFIFPCDAIASWNPFDRLRQMI  
 QKLLHGFKTKQVGENWCRFVEQFPPTLAERLITWYNI"

## ORIGIN

Query Match 80.0%; Score 16.8; DB 3; Length 2809;  
 Best Local Similarity 90.0%; Pred. No. 1.8e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGGCAGCGCCCAACAG 20  
 |||||  
 Db 507 GATTGGCAGCGCTCACAG 526

RESULT 12  
 BT003191

LOCUS 3377 bp mRNA linear INV 22-JAN-2003  
 DEFINITION Drosophila melanogaster RES9670 full insert cDNA.  
 ACCESSION BT003191  
 VERSION BT003191.1 GI:27819794  
 KEYWORDS FLI\_CDNA.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 3377)

AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
 Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,  
 George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G.,  
 Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,  
 Patel,S., Phouantenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.  
 and Celisner,S.

TITLE Direct Submission  
 JOURNAL Submitted (22-JAN-2003) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA

## COMMENT

Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720  
 This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our web site  
 at <http://fruitfly.berkeley.edu> or send email to  
[cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

## source

1. 3377  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="Y; cn bw sp"  
 /db\_xref="taxon:7227"  
 1. 3377  
 /gene="Trn"  
 /db\_xref="FLYBASE:FBgn0024921"  
 147..2828  
 /gene="Trn"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RES9670p"  
 /protein\_id="AAO24946.1"  
 /db\_xref="GI:27819795"  
 /db\_xref="FLYBASE:FBgn0024921"  
 /translation="MTWEPQEGEQIIAILKESQSDPTATQNAVMKLEEFNRYPDF  
 NNLYIYVLTKEDEFTSLGSLIKNNIRHGTTLQPEIYEIKHECLQAVGSSP  
 LIRATVGLITTIASNGLNHPQLPSLCLMDNQDYNVCEGAFSALQKICEDSAEI  
 LDSAALNPLNIPKFLYFKHSPKIRSHAIACINQFIINRSQALMLNIDSFIEL  
 FHLSSDEHVRNGLVGLVLLKVRMDLPHMSQIIEYMLLRTDQDGVLEBASE  
 FWLSLAQCTKDVLAAPYLAQAPLVARGMYSEVDIILLKGNVEDDMVPREDIR  
 PREKSRATHIRSTQEGGAGATGDDDDFEFGMDSSLSSENLKCSAAALDVLAN  
 VFREDCLPIVPLIKETLPHQEWIKESGVALGALAEQCMQOHLPFLIPYLISC  
 LSDKALVRISITLUSRVANVWVQPHDQYILKPLMEKLLKILDSNRKRVQEAACSF  
 ATLEEACETLVPYLYIKTLVFAFSKYQHNKLLLYDAVGLSDSVHGLNKRPQYI  
 DIIMLEPIIDKDDKDLFPLECLSLIATFALQSGFLPCDPYRRCILIRQTI  
 NQMLCKQNTYDHPKERMIVALDLSLGAELDRHIELTVLGNNSIMHLLYQCMQDV  
 LPVROSSFALLGLTKACFPVHPFMADFFPIILGONLNPDISVCNNATWALGICM  
 KLGEEFKOYRLVLSOLFILNRPNTPKLLNTATIGELGVCEVAVPYLPEFVR  
 QWCTSLRHINDNDEKSAFRGCMCHMTIVNPAGVADFIFPCDAIASWNPFDRLRQMI  
 QKLLHGFKTKQVGENWCRFVEQFPPTLAERLITWYNI"

PHLSDEDEHVRKVNCHGLVLMLEVRMDLPHMSQILLEYMLRTQDDEGVLEASE  
FMLSLEQSI CKDVLAPYLAQAPVLVRGMRYSVDVILLKKNVEEDMDVPRDIEDIR  
PHFKRATHTIRQGGAGATGDDDDDEFDDDDSSLSWNLKCSARALDVIAN  
VREDCLPVVILIKETIRYHQBWIKESGVLGAIAECMQMIOHLPILIPYLISC  
LSDKKALRSITCWTLSRVANWVQPHDOYLKPLMEELIKILDSNRVQBAACSAF  
ATLEBACTVLEPYLILKTLVFAFSKYQHKNLILYDAVGTLDASVGHNLKPYOI  
DILMPILDKWLLKDDKDLPLLECLSIATLQSGFLPYCDPVRCTSLIQCOTV  
NOEMLCQKQYDHPDKEMIVALLDLISGLAELDRHETLVANSIMHLLYQCMQDV  
LPEVRSSFALLGDLTKACFPNHPFMAFFPILGONLNPDPISVCNATWALGICM  
KLGEETKQIRLVLSDFIIRKNTPKTLTALNTLGRIGVYCFVEVAPYLPFVR  
QWCTSRHNDNDKDSAPFGCHMITVNPAGVADFFPFCDAIASWNPDPDLHQM  
QKILGHFKTQVGEENRRFVEQFPPTLAERLTMTYNI

ORIGIN

Query Match 80.0%; Score 16.8; DB 3; Length 3377;  
Best Local Similarity 90.0%; Pred. No. 1.8e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GATTGGCGAGCGCCCAACAG 20  
|||||  
Db 630 GATTCCGACGCTCAACAG 649

RESULT 13  
AE004465/c  
LOCUS  
DEFINITION  
Pseudomonas aeruginosa PA01, section 26 of 529 of the complete  
genome.  
ACCESSION  
AE004465 AE004091  
VERSION  
AE004465.1 GI:9946107  
KEYWORDS  
SOURCE

ORGANISM  
Pseudomonas aeruginosa PA01  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE  
1 (bases 1 to 11259)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S.L., Runagle, W.O., Kowalik, D.J.,  
Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,  
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,  
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
Wong, G.K.-S., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen  
NATURE 406 (6799), 959-964 (2000)  
20437337  
10984043

REFERENCE  
AUTHORS  
2 (bases 1 to 11259)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S.L., Runagle, W.O., Kowalik, D.J.,  
Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,  
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,  
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H.,  
Hancock, R.E.W., Lory, S. and Olson, M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University Of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 11259)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)  
Direct Submission  
Submitted (04-FEB-2003) Department of Molecular Biology and  
Biochemistry, Simon Fraser University, 888 University Dr.,  
Burnaby, British Columbia V5A 1S6, Canada

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
CONSTRM  
TITLE  
JOURNAL  
Pseudomonas aeruginosa PA01 genome annotation,  
updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,  
from PseudoCAP (see <http://www.pseudomonas.com> for latest updates  
and links to alternate annotations). PseudoCAP is coordinated by  
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert  
E.W. Hancock (University of British Columbia, Canada). We welcome

COMMENT  
This represents the February 3, 2003 version of the continually  
updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,  
from PseudoCAP (see <http://www.pseudomonas.com> for latest updates  
and links to alternate annotations). PseudoCAP is coordinated by  
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert  
E.W. Hancock (University of British Columbia, Canada). We welcome

submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.  
'Protein name confidence' is used to rate our confidence of the  
accuracy of the protein name.  
Class 1: Function experimentally demonstrated in P. aeruginosa.  
Class 2: Function of highly similar gene experimentally  
demonstrated in another organism (and gene context consistent  
in terms of pathways its involved in, if known).  
Class 3: Function proposed based on presence of conserved amino  
acid motif, structural feature or limited sequence similarity  
to an experimentally studied gene.  
Class 4: Homologs of previously reported genes of unknown function,  
or no similarity to any previously reported sequences.

FEATURES

Location/Qualifiers  
1. 11259  
/organism="Pseudomonas aeruginosa PA01"  
/mol\_type="genomic DNA"  
/strain="PA01"  
/db\_xref="taxon:208964"  
153..1433  
/gene="gabT"  
/locus\_tag="PA0266"  
153..1433  
/gene="gabT"  
/locus\_tag="PA0266"  
/note="Protein name confidence: Class 2 (high similarity  
to functionally studied protein); Alt protein name: gaba  
transaminase"  
/codon\_start=1  
/transl\_table=11  
/product="4-aminobutyrate aminotransferase"  
/protein\_id="AAG03655.1"  
/db\_xref="GI:9946108"  
/translation="MSKTNESILKRRQAAPRGVQIHPVVAERASTVWDVEGREY  
IDFGAIANTLHLPKVI AAVOSLQKLSHTCFVLAYEPIELAEIAKRVGDF  
PKKTLVTSGEAVENAVKIAAATGRAGVIAFTGAYHGRMTLGLTKVVPVSAGM  
GLMPCGIFRALAPLHVLVSDEDSIASIERIFKQDAQODIAAIIIEVQGGGPFVN  
SKFMQRALCDQHGILLIADDEVOTGAGRTGTFFATLQGLIVPDLITTFASVGGGFP  
ISVAGKAIMDAIPAGLGSTYAGSPIACAAALAVLKVFEEKLLERSQAVGERLKA  
GIRETQAKHVGIDVRGLGSMVAIEFEGDTHKPAELVSKIVVRAEKGILLILSCG  
TVYVIRLMPVTTPDAQLEKGLAILAEFCDELA"  
1759..2958  
/locus\_tag="PA0267"  
1759..2958  
/locus\_tag="PA0267"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03656.1"  
/db\_xref="GI:9946109"  
/translation="MSAAKKAFLIADPDPMSRDLGLQVLGVRCARLVLCGGGGE  
ALAHCRRRFALILAEINLPVDGPELLREARLRSVAEPFLLISDRADQASVRAAV  
ALAPAYLVKPFQAEINLQRLGKGVVACPLPERDAGENLEAFERARDSAEG  
APLLADVAASDRCLSAEHLPLVLEAFGRDPQITAGLIAANSAAHVGPACOTLGA  
QALRUGHSNLNVLFSLQRAIQGLFELIAERAMHPWDLISQRCADIWELADALGA  
DVERCYTAGLLRLGLDALLRTLQDWCDDGALDEARLDLGRGFGSALRAWR  
LPLERLRLAAAYQFGVLSREELILSLATQAASLPEDDAEQALSSKAARMLGLDGER  
LAKLIQP"  
complement(3112..4533)  
/locus\_tag="PA0268"  
complement(3112..4533)  
/locus\_tag="PA0268"  
/note="Protein name confidence: Class 3 (function proposed  
based on presence of conserved amino acid motif,  
structural feature or limited sequence similarity to an  
experimentally studied gene)"  
/codon\_start=1  
/transl\_table=11  
/product="probable transcriptional regulator"

protein id="AAG03657.1"  
/db\_xref="GI:9946110"  
/translation="MELQLILDGSHDLVGOLYRQLRDAIILHGRILATGDRLPATRYLAE  
FLEYSRTVSAYERLASEGLEGRVSGTFVARGPAGEAPQORADGSLPAPARWARI  
EPQGLMDQAPLRISFAGTIDKRRFFDTWTCVQLQRLQORSFAFYKETEEOPE  
LRLAISRYLAIRAVQCDNMALMTQGAQQLDLARMLDPCGLAVVEPGYPFARA  
SLLATGAEVRGPVDAEGRLVDLFDNAKLYVTFPSHQFPLGMPMSLERLQLLDWAQ  
RRLALITDDYDSYRDRGSLRSLRSDRNGQVAYLGTFSKTLFPLRLGLYLLPAS  
LREAVLAARLVDRSHSTLQORALALFIQORDPARHTRRLQREVAQRGRGLRLRHGD  
LSPLEALPAVAGIHLAARCEPDVLRARIARHLRQQGLEVGTLPFPFAETPAAGTFLF  
GFGIERDDIDPDLRLAELRR"  
4660..5097  
/locus\_tag="PA0269"  
4660..5097  
/locus\_tag="PA0269"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAG03658.1"  
/db\_xref="GI:9946111"  
/translation="MTRLEWAKASPDYAAMLGLEKALAKAGLERPLIELVLRISQ  
INGCAYCNMEHANGARAGATQRLQALCVNQETPYTPPRERAALANTQLARLSQGA  
LPHGLDELREHFDKIEBELTLAVSAINAWNRFGVMGMOPE"  
5109..5516  
/locus\_tag="PA0270"  
5109..5516  
/locus\_tag="PA0270"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03659.1"  
/db\_xref="GI:9946112"  
/translation="MRANRSIACLLLAALPTLSAHVAEKVETLHPQALAEITPHGRGL  
MLTVSPKFGQVSAPHVPGSVFVYVLEGVTQOLEGGSVETVVGQSVYEPANVPPLL  
ARNPDMKPAKLLVWELKEGDPILKPLERGAH"  
5530..5834  
/locus\_tag="PA0271"  
5530..5834  
/locus\_tag="PA0271"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG03660.1"  
/db\_xref="GI:9946113"  
/translation="MSASQVIRIRANGOSERLGLLQQLVFPGLAPCGLYQVRRD  
ALRPDLWVCSQWREARQFDFYLAGGIQVFAEVLWRSLSAASLEVQDDAG"  
complement(5831..6763)  
/locus\_tag="PA0272"  
complement(5831..6763)  
/locus\_tag="PA0272"  
/note="Protein name confidence: Class 3 (function proposed  
based on presence of conserved amino acid motif,  
structural feature or limited sequence similarity to an  
experimentally studied gene); Subcellular localization:  
Cytoplasmic; Subcellular localization confidence: Class 2"  
/codon\_start=1  
/transl\_table=11  
/product="probable transcriptional regulator"  
/protein\_id="AAG03661.1"  
/db\_xref="GI:9946114"  
/translation="MLIANGLRKVMQDILLVFLQVHQRNLVEVAHLNLSPTVSVC  
LKLROAFDELFIATRGQMOPHKSRAMLPHVEMLARINACHAAGSFDPRGPRT  
FTLCAPEYVELLILPALLRLAANGPRICLDIRRLGPPELTALDGOVLDVDPGE  
YHRPAPLQALALFSDLLCVQDRAPQERHLELEDEFCCRRTVFTPLDASANLVDT

WLRROGRSREVAARANSYVALQLLSGSDFVLMPLPRRIHALLDGRHHACLAPCLPP  
FALHLSWRAARQDPANLWLRBOVLIVCAELGLL"  
complement(6813..8027)  
/locus\_tag="PA0273"  
complement(6813..8027)  
/locus\_tag="PA0273"  
Query Match 80.0%; Score 16.8; DB 1; Length 11259;  
Best Local Similarity 90.0%; Pred.No. 1.7e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 ATTGGCAGCGCCCAACAGC 21  
|||||  
Db 10089 ATTGGCAGCGCCCAACACC 10070  
RESULT 14  
AF521878 17665 bp DNA linear BCT 29-JUL-2002  
LOCUS  
DEFINITION Streptomyces narbonensis desosamine biosynthetic gene cluster,  
partial sequence.  
ACCESSION AF521878  
VERSION AF521878.1 GI:22001230  
KEYWORDS Streptomyces narbonensis  
SOURCE Streptomyces narbonensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycetaceae; Streptomycetaceae; Streptomycetes.  
REFERENCE 1 (bases 1 to 17665)  
AUTHORS Bate,N. and Cundliffe,E.  
TITLE The desosamine biosynthetic cluster of Streptomyces narbonensis,  
producer of narbomycin  
JOURNAL Unpublished  
AUTHORS Bate,N. and Cundliffe,E.  
TITLE Direct Submission  
JOURNAL Submitted (14-JUN-2002) Biochemistry, Leicester University,  
University Road, Leicester LE1 7RH, UK  
FEATURES  
source  
1..17665  
/organism="Streptomyces narbonensis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:67333"  
<1..528  
/gene="nbmA"  
<1..528  
/gene="nbmA"  
/note="terminal thioesterase I (TEI) domain of type I  
polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="NbmA"  
/protein\_id="AAM88350.1"  
/db\_xref="GI:22001231"  
/translation="AVLRAAGAPVLLGHSGGALLAHLAFRLERAGHTAPAGIVLV  
DPYPGHQEPLEKVSROLGRLGELPMSDARLLAMGRVAFRLAGPRGRSSAPVL  
LVRAERLGDMPEREGDWRAHNDLPHSVADVPDGHFTMRDHAPAVAVAVLWLDGIE  
RDTAAGRTRTQAGADQ"  
525..1370  
/gene="nbmB"  
525..1370  
/gene="nbmB"  
/note="thioesterase II (TEII)"  
/codon\_start=1  
/transl\_table=11  
/product="Nbmb"  
/protein\_id="AAM88351.1"  
/db\_xref="GI:22001232"  
/translation="MTDRVLNVDSLSLWIRRHPSNSAVRLVCLPHAGGSASYFFRFS  
EELHPSVEALSVOYPGQDRARPCLVESBELAEHVVAATEAWREGRLAFFGHSILGA  
SVAYEARAILIQHGVPRPEGLYVSGRAPSLAPDRLVHOLDRAFLAEIPQLSGTGA  
FLQDDELRLVLPALRSDYKAAITYEHRPSAKLACPMALAGDRDPKAFLENEVAENWR  
HTSGPFCLIRASQGHFYLDNQWHGICNDISDHLVLTNGADPSRVVQPTTKLIEGAAR

```

WONPR"
1367. .2713
/gene="nbc"
1367. .2713
/gene="nbc"
/notes="putative NDP-deoxyhexose 3,4-isomerase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88352.1"
/db_xref="GI:22001233"
/translation="MTDDPARALTQPLGRVTSAVADPELGAHLLETRGIHWHANG
DPYAAVLRCQADDPYAYERVRARGVLSFPTGSWVTAADHVAASVLCSTEFVPCAD
SVFPOQVLSYGECCLEHQAPRAAAEVPAGAWQORRAVVEVERHQLLEGAPDSRA
SVAFELAGVGPVAVTAAAVLGVPGAGRGFAELVERLPLSDLSLAQSLRVTRA
ACGALAALFALLAAGVSAATGQSDPSGSPASPPSDSSSGPTAPSGPTAPSGV
AGRPDALLSALGVTAAVOLTGNVAVALLAHPEQRELCARGLGAAAVEETLRDPP
VQDRAVARVETELAGRLPAGHVVVLTAGTGRDGVFTDPERFDLAPDAIHAL
HPAGPYGPVASILVRLQAEVALRTLAGRFFGLRKAGVLRPRRAPVGRGSLFVSGS
2782. .4062
/gene="nbc"
2782. .4062
/gene="nbc"
/notes="desosaminyl transferase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88353.1"
/db_xref="GI:22001234"
/translation="MRVILTSFAHHTHYVGLVPLAWALLAAGHEVVASQALTDIT
GSLAAPVGTGDLHIEYVRMAGEPRPHPAIFAGASGEFLDWDHALGIDALLAPY
FYLLANDNSMIDLVGFARSQDLVWSPFTYAGAAQVTAAGAAHAWLWPGDVWGS
ARKFVALRDRQPEHREDTAEWLTWLDGFGAFEBELLTGQFTVTPPSLRDLTA
GLTQVGVYVNGTSVVDMLAEPARPVRVCLTLGVSAREVLGSGVQADILEALA
DLDELVAITDAGCAEIRSPKHTFTDVPVHALLPSCSAIVHHCAGTATAVIN
AVQVLMELWDAVPKARAVAGQAGFLPPAGLTQAVRDADVRLDDPSVLAAGR
LRETHRDPTAGIVPELERLAAQHRPPADARR"
4151. .4864
/gene="nbc"
4151. .4864
/gene="nbc"
/notes="N-methyltransferase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88354.1"
/db_xref="GI:22001235"
/translation="MYEVDHADVVDLFLYLRGKDYAAEASDIADLVTRTRTPASSLLD
VAGTGTGHLSEHFKEGDGTAGLELSEDMLTARKRLPDATLHOGDMRDRLGRFSAV
VSPFSGVYLRTTAEALDAVASAAHLEPGGVVVEWPPETFADGWVSADVVRDGG
RTVARVSHVRDGDATKRMVEHFTVADPGRGVRHFSVHLITLFHRAEYEAFTAGLR
VEYLEGGSPGRGLFVGVA"
4908. .7424
/gene="nbc"
4908. .7424
/gene="nbc"
/notes="beta glucosidase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88355.1"
/db_xref="GI:22001236"
/translation="MHOAQREKSTVCKTRI PVRRSRTTTPRAFTLAVGTLGAGTTV
AAAPAGAGTGHVYTSKAELVAQMLTDEKISFVHWALDPDQNVGLYGLPRLGIP
ELRAADPGLNGRLVGRTAELVAQMLTDEKISFVHWALDPDQNVGLYGLPRLGIP
MMNIRVPHGRNYETSEDPVSSRTAAVAQIKIGIQAGLMTAKHFAANNQNNRPS
VNAVTDEPQTLRETFEASAKAGASFCAYNGVNGKSCGNDELLNNVLTQWGF
QGWNSDWLATRGDITKGLDQEMGVLEPCDI PPGEPPSPFAGDALQVAVLNGTV
PEAAVTSABRI VNMQKFGILLATPAPRPERDKAGACQASRVKVAENGAVLNEGA
LP LAGDAGKSTAVLGPVAPKVTGLGSAHVDPDAAPIDITIKARAGATVYETG
EETGTIRPAQQLSPAFNQHQLEPFGKAGALYDGLTVPADGGEYRISVKATGYATVQ
LGSHTIEAGQVYGVSSVPLKLTGTHKUTISGFAMGATPLSLGELGVNVTPEADATIA
KAVESARKAKTAIVFAYDDGTEGVDRPNLSLPGTQDKLI SAVADANENTIVILNTGSS

```

```

VLPWLSKTRAVLDMWYPGQAGABATAALLYGVDPNPSGKLTQSFPAAENOHAVAGDPN
RYPGVDNQOTYSEGHVGVYKVPFLPFHGLSYTSFTQSAPTVVTRTSTGGLK
VTVTRNSGORAGOEYVQAVLGASPKVTPAQAEKLVGYTKVALAGESKTVTVNVDK
RQLQYWDAAASDSWRTGTGSLLOTGSSSADLKNSQVTVW"
complement (7514. .8653)
/gene="nbc"
complement (7514. .8653)
/gene="nbc"
/notes="aminotransferase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88356.1"
/db_xref="GI:22001237"
/translation="MTARTETGRVPFLDLKAAVEELRPETDAAIARVLDGSRVYLGP
LEATPEAFATCYCTEDTHAVGVNSMDALQALRGLGIGGDEVI VPSHYTASWLAVSA
TGATPVPERNEHPTDPLLVKAIPTPTALLPVHLYGHPADLALDRDLRHLGH
LVEADAQAHGAYEGRIGAGASVAFSPYKGLGCGGGA VVTGDPAALAEIRML
ANYGRQYSHETKGTNSRLDEIOAAVLRIRLAHLDSWNGRSTLAAEYLSGLAGLPG
IALVPTAPGTFDVFVWHLFTVTRDRDELRLDSRGVDLTHYFPVPHLSPAYAQAPP
EGSLPRAESFARQVLSLPIGPHLERTQALRVIDAVREWAEOGRP"
complement (8650. .9684)
/gene="nbc"
complement (8650. .9684)
/gene="nbc"
/notes="NDP-glucose 4,6 dehydratase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88357.1"
/db_xref="GI:22001238"
/translation="MRLLVTGAGFGISGHFVRLQAGAYPDVPAGEVIVLDSLTYAGN
RANLAVPADPRFLRFVGHGDIRDADALLARELGVDAVHFAESHVDSIAGASVFTGT
NQVQTLLQCALDAGVRVHVSTDEYVIGSDGWNESFLEPNSPYAAKSGASDL
VARAYRTYGLDVITRCCNNYQVHPEKLIPLFVNLDDGTLPLYDGGANVREW
HTDHCRTGIALVLAGRAGEVHYGGRELTNRLTGILLDSLGADWSVVRVADROG
HDLRYSLDGGKIERELGYRQIATADGLARTIRWYRENDRWWEPLKATAPQPPATAPQ
PPATAVGVSA"
complement (9759. .10637)
/gene="nbc"
complement (9759. .10637)
/gene="nbc"
/notes="NDP-glucose synthase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88358.1"
/db_xref="GI:22001239"
/translation="MKGIVLAGSGSOTRLHPATSVISKQILPVNPKMIYPLSVMLG
GIREQIISTPOHIELFQSLNCRHLGIELDYAVQKEPAGIADALLVAGHIGDDTC
ALILGDNILFHGGLYLLRDSIARLDGCVLFGYPVKDPERYGAEDVATGRITGLVEK
PEKPSHLAVTGLYLDNDVVDIAKNIIRPSRGELEITDVRVYLERGRAELVDLGRG
FAWLDTGTHDSLLRAAQVQVLEERQGVNIAGLELIAFRMGFIDAEACHSLGELSR
EYGSYLMETIAGREGAP"
complement (10840. .12297)
/gene="nbc"
complement (10840. .12297)
/gene="nbc"
/notes="putative NDP-deoxyhexose 3,4-reductase"
/codon_start=1
/transl_table=11
/product="Nbc"
/protein_id="AAM88359.1"
/db_xref="GI:22001240"
/translation="MTAPAVATAPSAARCAHPGADLGAARAVAGQSAAAGGLVPDSEA
GSTARHLVGLAVYNGSPFTPLEEARHDLGVDRDAFRLLALFPRPELRAVEGPA
GAYWNTLLFLERRGVFSALAGKRVFFYSVGLYFGPCMFCHFCVRYTGARYDPSA
LDTGAMNFSDIDEIPAGNHALYFSGLEPNTNGLTAAHAAHGRURFTVYINSF
ALTERLERQGLWGLHVAVTSLYLNDDEYEYTTGKKAARVRRENURRFOOLRAER
EAPINLQAYIVLPGRAHRLDLVFIADLDESQGTIDFVNIREDYSGEDGKLQ
ERRAELQALAAFEERVERTPGLHIDYALNSLKAGADALLRIKATMRPTAHQ
VAQVLDLGDVLYREAGFPDLDGATRIYAGRVGCTSLTEVVRDVERGGRVAARDG

```

Query Match 80.0%; Score 16.8; DB 1; Length 17665;  
Best Local Similarity 90.0%; Pred. No. 1.6e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTGCGCAGCGCCCAACAGC 21  
| | | | | | | | | | | | | | | | | | | | | |  
DB 15002 ACTGCGCAGCGCCCAACAGC 15021

RESULT 15  
AC013151  
LOCUS 21788 bp DNA linear HTG 03-NOV-1999  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
ACCESSION AC013151  
VERSION AC013151.1 GI:6223180  
KEYWORDS HTG; HTGS\_PHASE2  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 21788)  
Adams, M. and Venter, J.C.  
Direct Submission  
Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDM:10213745 by the submitter.  
For further information on this sequence you may e-mail to  
fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1..21788  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"

ORIGIN

Query Match 80.0%; Score 16.8; DB 2; Length 21788;  
Best Local Similarity 90.0%; Pred. No. 1.6e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTGCGCAGCGCCCAACAGC 21  
| | | | | | | | | | | | | | | | | | | | | |  
DB 6251 ATTGCTCAGAGCCCAACAGC 6270

Search completed: April 29, 2004, 06:01:43  
Job time : 321.194 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 362.165 Seconds  
(without alignments)  
2872.264 Million cell updates/sec

Title: US-10-624-714-7

Perfect score: 24

Sequence: 1 ggagcgcttggtacggtcgcatc 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.mu.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	100.0	1097	1	AF074390	AF074390 Mycobacte
2	24	100.0	1097	6	A67974	A67974 Sequence 19
3	24	100.0	15670	1	AE006951	AE006951 Mycobacte
4	24	100.0	86114	6	AX704277	AX704277 Sequence
5	24	100.0	324050	1	BX248335	BX248335 Mycobacte
6	24	100.0	342416	1	EX842573	EX842573 Mycobacte
C 7	19.8	82.5	110000	2	AC112426_0	AC112426 Rattus no
8	19.8	82.5	183433	2	AC142436	AC142436 Rattus no
9	19.8	82.5	216124	2	AC098305	AC098305 Rattus no
10	19.8	82.5	276158	2	AC129229	AC129229 Rattus no
C 11	19.2	80.0	154038	8	NCB8G12	BX294027 Neurospor
C 12	19.2	80.0	308950	1	MLEPRTN9	AL5831925 Mycobacte
C 13	18.8	78.3	174622	10	AC079436	AC079436 Rattus no
C 14	18.8	78.3	232195	2	AC106493	AC106493 Rattus no
C 15	18.4	76.7	134855	2	AP004077	AP004077 Oryza sat
C 16	18.4	76.7	143943	8	AP004997	AP004997 Oryza sat
C 17	18.2	75.8	113193	1	AF357202	AF357202 Streptomy
C 18	18.2	75.8	113193	6	AX703543	AX703543 Sequence
C 19	18.2	75.8	164493	2	AC118618	AC118618 Mus muscu
C 20	18.2	75.8	185795	10	AC127333	AC127333 Mus muscu
C 21	18.2	75.8	194309	10	AC125213	AC125213 Mus muscu
C 22	18.2	75.8	234149	2	AC139055	AC139055 Mus muscu
C 23	18.2	75.8	252988	2	AC105511	AC105511 Rattus no
C 24	18.2	75.8	257145	2	AC122573	AC122573 Rattus no
C 25	18.2	75.8	281951	2	AC110093	AC110093 Rattus no
C 26	18.2	75.8	283356	2	AC109765	AC109765 Rattus no
C 27	18.2	75.8	291394	2	AC113712	AC113712 Rattus no
C 28	18.2	75.8	298698	2	AC094778	AC094778 Rattus no
C 29	18.2	75.8	328905	2	AC094646	AC094646 Rattus no
C 30	17.8	74.2	3204	8	AK101077	AK101077 Oryza sat
C 31	17.8	74.2	11208	1	HHRNAP03	X57144 H. halobium
C 32	17.8	74.2	11894	1	AE005139	AE005139 Halobacte
C 33	17.8	74.2	20468	1	AE008840	AE008840 Salmonell
C 34	17.8	74.2	134900	8	AP005414	AP005414 Oryza sat
C 35	17.8	74.2	136420	2	AC104846	AC104846 Oryza sat
C 36	17.8	74.2	220220	2	AC125657	AC125657 Rattus no
C 37	17.8	74.2	292100	1	SC0939121	AL939121 Streptomy
C 38	17.8	74.2	292200	1	SC0939129	AL939129 Streptomy
C 39	17.8	74.2	301950	1	AP006570	AP006570 Gloebact
C 40	17.8	74.2	318200	1	MLEPRTN6	AL583922 Mycobacte
C 41	17.8	74.2	348517	1	BX248354	BX248354 Corynebac
C 42	17.8	74.2	348970	2	AC138608	AC138608 Oryza sat
C 43	17.6	73.3	573	1	AF526299	AF526299 Mycobacte
C 44	17.6	73.3	1284	1	AY077756	AY077756 Streptomy
C 45	17.6	73.3	2064	10	AF441120	AF441120 Mus muscu

ALIGNMENTS

RESULT 1  
AF074390  
LOCUS AF074390  
DEFINITION Mycobacterium tuberculosis heparin-binding hemagglutinin (hbha)  
ACCSSION AF074390  
VERSION AF074390.1 GI:3309265  
KEYWORDS Mycobacterium tuberculosis  
SOURCE Mycobacterium tuberculosis  
ORGANISM Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 1097)  
REFERENCE  
AUTHORS Menozzi, F.D., Rouse, J.H., Alavi, M., Laude-Sharp, M., Muller, J.,

Pred. No. is the number of results predicted by chance to have a





```

/translation="MPDAGAGSLRSWAVALRTNPPADGPTDTVTRMLVTVTRAAVLV
MTLVSGVAGLGAIGFGLDRWLVLWVSHAPIANNLMLNDLYDDVGTDSATYARA
RYACHPAATGANAAVYTPRRRTSCGSPRALEPTTPRWARAVGRSCWRSPGTCAP
RC"
complement(971. .1741)
/gene="MT0489"
CDS
complement(971. .1741)
/gene="MT0489"
/notes="similar to GP:3063879; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, TetR family"
/protein_id="AAK44712.1"
/db_xref="GI:13879996"
/translation="MCQDPQLAVSGASSEWELATVAERIPAVTVKTGKRKRWHQH
KVRNRNLDCTTEAIRHGRFLMSIEAIEIGVSKTVLYRFYDKNDLTTAVMMRFT
OTTLLPNXADGADMGFELTIRVYVETVAOPEPYRFVWANSASKSVIADS
ERIALMLVLRMRMOEAGMTGGVEPMAYLVGGVOLATHSWMSDPRMSSDELIDY
LTMLSWALGCGIVGAGSLKFRQHPHFPIVPAWGV"
1956. .3182
/gene="MT0490"
CDS
1956. .3182
/gene="MT0490"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44713.1"
/db_xref="GI:13879997"
/translation="MALRMKVVAVGFLVGATGVFLACRWAGADGADHAWLGYLGAA
EAGWGLADMFVATLKFPLGPIPIHTAI IKKKDOLGELGTFRVNFLLPFPVVE
TKLRDAOI PSLKWLSEAHAORVAETATLVRLVLLERDDEIOQVDMILVRRIA
EPQWGPAGRVLATLLENQGEAFIOLLADRAFQWSINAGVVCQVVERDSPNSRF
IDHLVGRIRHELMETDKVNRPNHELRSAFLDFPADDLQHPDPATVADAIKE
ELWARDIATAAAWTKLRLVLEGVDDPSALRTRITDAVIRIGESLRDDADLRDK
VDSWTVRAAQLHVSEYGVETAITITETIERWDAEASRRIELHVGRDLQPIRINGTV
GAVAGLAIYAQLLP"
3269. .3691
/gene="MT0491"
CDS
3269. .3691
/gene="MT0491"
/notes="similar to GB:X66407 PID:46399; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, PbxX family"
/protein_id="AAK44714.1"
/db_xref="GI:13879998"
/translation="MSSEKLAAKVSTKASVDSDIGSFIRSQRETAHVSMQLAERS
GVSNPYLSQVERGKPSADVLISQIAKALRVSAEVLVYRAGILEPSETSQVRDAIITD
TAITEROKOILLDIYAGFTHONETRECEPDPPTDD"
complement(3730. .3867)
/gene="MT0492"
CDS
complement(3730. .3867)
/gene="MT0492"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44715.1"
/db_xref="GI:13879999"
/translation="MSGPADPVTTCLICTSSTESLSSAADRDLQPAVDLTQMTVLQ
V"
4030. .4644
/gene="MT0493"
CDS
4030. .4644
/gene="MT0493"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44716.1"

```

```

/db_xref="GI:13880000"
/translation="MKGIPMAENSNIDDIKAPLLAALGAADLALATNELITNLBERRA
BETRTDTSRVEESRARTLKQEDLPEQTELREKFTAEELRKAAGYLEAATSRINE
LVERGEALERLSQSPFEYSARAEYVDQAVELTQELGTVASQTRAVGERAAKLV
GIELPKAAAPAKKAPAKKAAAPAKKAAAKAPAKKAAAKKVTQK"
complement(4657. .4809)
/gene="MT0494"
CDS
complement(4657. .4809)
/gene="MT0494"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44717.1"
/db_xref="GI:13880001"
/translation="MHRVHRCHQDQDOYQHDGTHEVAHAFRLCGLHPTRLRLVAPV
ATRSRW"
4925. .5470
/gene="MT0495"
CDS
4925. .5470
/gene="MT0495"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44718.1"
/db_xref="GI:13880002"
/translation="MCSGWECPPVPACIWTSCGSPFRFRASRANGMKALVAVSAVA
VVALGVSAQADPADPCAGEANYGGPSSRLVDHTEWAGWSLPSLVYVPSQVGR
TASRLGVWAAADAAEVLALUSPEADTAGMRAQFICHWQYAEIROFGKPSMNLFWRP
VDDSEMLASGNCNPGSPESF"
5470. .6144
/gene="MT0496"
CDS
5470. .6144
/gene="MT0496"
/notes="similar to SP:P39121 PID:809660 PID:1408506
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="deoxyribose-phosphate aldolase"
/protein_id="AAK44719.1"
/db_xref="GI:13880003"
/translation="MLQPTRAQLAALVDHTLLKPTTRADVAALVAEAEELGVYAVC
VSPKVPVAVQGVRAAVGAVTGPSPGKHVSVAHEAAALASGEIDMDVIDIGAL
CGDIDAVRSDIEAVRAAGAVLKIVESAVLLGQSNATLVDACRAEDAGADFVKY
STGCHPAGGATVRAVELMAETVGPRLGVKASGIRTAADAVAMLNAGATRLGSLGTRA
VLDGLS"
complement(6169. .7215)
/gene="MT0497"
CDS
complement(6169. .7215)
/gene="MT0497"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44720.1"
/db_xref="GI:13880004"
/translation="MTNPGPNPDSPWARPQDQGLARPPASSEASTGRLRPGEPAAG
HIQEPVSPPTQEPQFPQTEHLAASHAHTSRGQRAHQAWDPTGLAAQEEPPAAVKT
KERARDPLTVLFTLVISVLVAGLIGGELYARHVANSKVAQAVCVVQDQATASFG
VAPLLILVQVATRHFTNISVETAGNOIRDAKGMQIKLTIONVLKNTPNRGITGALDA
TWTSSSEGIKESVQNAIPILGAFVTVSSVTHPADGTVELKGLNNITAKPIVAGKGLE
LQIINFNTGFLSPKRETQVSTVSTNEFTSSITKNYPLGIHADSVQVTVSTGVVSPSTEDA
AIPGTQNPFCFSHI"
complement(7212. .8234)
/gene="MT0498"
CDS
complement(7212. .8234)
/gene="MT0498"

```

Query Match 100.0%; Score 24; DB 1; Length 15670;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCATC 24  
|||||  
Db 4446 GGAGCGTTGGTACGGTCGCATC 4469

RESULT 4  
LOCUS AX704277 86114 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 648 from Patent WO20074903.  
ACCESSION AX704277  
VERSION AX704277.1 GI:29538532  
KEYWORDS Mycobacterium tuberculosis  
SOURCE Mycobacterium tuberculosis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE 1  
AUTHORS Cole, S.  
TITLE Comparative mycobacterial genomics as a tool for identifying  
targets for the diagnosis, prophylaxis or treatment of  
mycobacterioses  
JOURNAL Patent: WO 02074903-A 648 26-SEP-2002;  
INSTITUT PASTEUR (FR)  
FEATURES  
source  
1..86114  
/organism="Mycobacterium tuberculosis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1773"  
/note="BAC-Rv265 deposited at the C.N.C.M. under the A.N.  
1-2628"

ORIGIN  
Query Match 100.0%; Score 24; DB 6; Length 86114;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCATC 24  
|||||  
Db 52795 GGAGCGTTGGTACGGTCGCATC 52818

RESULT 5  
LOCUS BX248335 324050 bp DNA linear BCT 11-JUN-2003  
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
2/14.  
ACCESSION BX248335  
VERSION BX248335.1 GI:31617046  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis subsp. bovis AF2122/97  
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.

REFERENCE 1  
AUTHORS Garnier, T., Eigmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,  
Pryor, M., Duchov, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,  
Harris, B., Akkin, R., Doggett, J., Mayes, R., Keating, L.,  
Wheeler, P., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and  
Hewinson, G.  
TITLE The complete genome sequence of Mycobacterium bovis  
ONLINE Publication  
REMARK PNAS 10.1073/pnas.1130426100 ( Microbiology )  
REFERENCE 2 (bases 1 to 324050)  
AUTHORS Garnier, T.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:tgarnier@pasteur.fr submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,

FEATURES  
source  
1..324050  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
132..2027  
/locus\_tag="Mb0290"  
132..2027  
/locus\_tag="Mb0290"  
/note="Mb0290", len: 631 aa. Equivalent to Rv0282, len:  
631 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.8% identity in 631 aa overlap). Conserved hypothetical  
protein, similar to Y14967|MLCB628.18c hypothetical  
protein from Mycobacterium leprae (573 aa), FASTA scores:  
opt: 916, E(): 0, (38.7% identity in 568 aa overlap). Also  
similar to Mycobacterium tuberculosis proteins e.g.  
Z94121|MTY15F10.26 (619 aa), FASTA scores: opt: 743, E():  
0, (29.9% identity in 612 aa overlap). Member of CFPO,  
CBXP family - 9 members in Mycobacterium tuberculosis,  
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD93154.1"  
/db\_xref="GI:31617047"  
/translation="MAGVGAGDSGGVERDDIGWVAASPVASRVNGKVADVVRPATC  
CRALGIATVQRKPPDLAARSQGFALTRVAHQDQCDATGLAAGDQSGIVLEASRT  
ATTAGVLQORVELADNALGFLYDTGLYLFRTATGDDDFHLAALASDTGGFEAFKA  
NVTSGITERRAGWRAWLAIVINRAERMSDVVKLTVMVNDPDLDEAFSHAAKIT  
LGTALRIGCMFAPALSYLEEDPGFVAVAVDGAALALVLRHVDDESASEVLODLVA  
APENEQVEALSDTSFGIVTTAGRIEMARLDPDPAPEGAEDFVDPAAHKAALL  
HEALQLAEFFGLDVKVQSRKSSVAMELVKRGITVAQTHHLVLEADGVLEDEA  
TIARVAKYCGLLKREINREVRADLIGCHIGETEAKTAKNIIDSLDGLVLEDEA  
YALVATGKNDPGLVAIDTLARMDNRDLVLIAGYRADDLKFDLNEGRGRFTR  
NIDFYSYTHSELVIAHKMAQRDSVFQSAHLHLEALFALAESEPTDNGIRSFTR  
DIAGNRFVRNIVERSEERFLDHSAGSGEFSDELMITADDVGRSVEPLLRG  
LGLSVRA"

2024..3640  
/locus\_tag="Mb0291"  
2024..3640  
/locus\_tag="Mb0291"  
/note="Mb0291", len: 538 aa. Equivalent to Rv0283, len:  
538 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 538 aa overlap). Possible conserved  
membrane protein, similar to several hypothetical  
mycobacterial proteins e.g. Z94121|MTY15F10.16|Rv3895C  
from Mycobacterium tuberculosis (495 aa), FASTA scores:  
opt: 698, E(): 0, (37.6% identity in 492 aa overlap);  
Rv1782; Rv3450c; Rv3869; and Y14967|MLCB628.16|MLCB28.17c  
from Mycobacterium leprae (481 aa), FASTA scores: opt:  
672, E(): 1.5e-31, (37.2% identity in 506 aa overlap).  
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
/codon\_start=1  
/evidence=experimental  
/transl\_table=11  
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"  
/protein\_id="CAD93155.1"  
/db\_xref="GI:31617048"  
/translation="MTNQCHDHPDHRSEFASRTPVNNNDPKVYRGGFVTRHQVTC  
WFFVMRIAGALHDEMLVDPARTOSRAVMGLVITLGLIGSFVFLRNGOG  
SNVLAADSTALYVRVGEQLHPVLNLSARLIVGRVSPVTVKSTELDQDFRGNLIG  
IPGAPRWQMTSDANMTVCDGLNAPSRGGAGGVTVIAGPLEDTPGARAALGPGQ  
AVLDVSGTLLMDGKRSPIDLDHVTSLGLGADVPAPRIIASGLNFIAPBPPL  
TAPIIPDAGNPAFGVPAPIGAVSSVAKDSKSTIDTVQYVAVLPDGLQISFVLA  
AIIIRNNNSYGIQPPRLGADGAVKLVSRVLDTRRYSPEVSLVDVDPDTQYVWSK  
PVGAATSLTLGASLPPVDAVHTVELVAGNGVATRVAAAGTGVFTOTVGGGPD  
APGAGSLFWSDTGVRYGIDNEPQGVAGGKAVEALGILNPPPIPWVSLFLVPGPT



TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
 COMMENT  
 On or before Nov 21, 2003 this sequence version replaced  
 gi:3261524, gi:3261526, gi:3261527, gi:3261529, gi:3242298,  
 gi:3261606, gi:3261689, gi:3261708, gi:3261703, gi:3261760,  
 gi:3261781, gi:3261797, gi:3261800, gi:3261825, gi:3261837.  
 Notes:  
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).  
 FEATURES  
 Location/Qualifiers  
 1..342416  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="Genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 223..2118  
 /locus\_tag="Rv0282"  
 223..2118  
 /locus\_tag="Rv0282"  
 /function="UNKNOWN"  
 /note="Rv0282, (MTV035.10), len: 631 aa. Conserved hypothetical protein, similar to Y14967|MLCB628.18c hypothetical protein from Mycobacterium leprae (573 aa), FASTA scores: opt: 916, E(): 0. (38.7% identity in 568 aa overlap). Also similar to Mycobacterium tuberculosis proteins e.g. Z94121|MTV15F10.26 (619 aa), FASTA scores: opt: 743, E(): 0. (29.9% identity in 612 aa overlap). Member of CFQX, CBXP family - 9 members in Mycobacterium tuberculosis. Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="CONSERVED HYPOTHETICAL PROTEIN"  
 /protein\_id="CAAL17357.1"  
 /db\_xref="GI:2909471"  
 /translation="MAGVGEQSGVERDDIGWVAFSVASRVNGKVDADWGRFATC CRAIGAVYQRKPPDLAAARSGPAALTRVAHDOCDANTGLAAGDOSIGVLAASST ATTAGLQORVELADNALGFYDTGLYLRFRATGPDFDFHLAYAAALASTGPFEFAXA NHVSGITERAGARAAKWLAVINRAERDSVVKLLTPWNPDPDDEAFSAAKIT LGTALRGMPALSYLEEDPGVAAVADGAKALVLRHVDESSASEVJODLYA AHPEQVEQALSTSFQVITTAGTTEARTDPWDPAEPGAEFVDPFAAHERKAALL HEALQIAFTGLDEVKRQVRSLSVAMELVKQRLTVAQRTHLVFAFPPTGKLT TIARVAKYCGLLGKREINREVRADLIGHIGETEAKTNAIDGALDGLVFLDEA VALAVAKNDGIVADITLLARNENDRDLVIIAGYRADLDKFLDTNGLSRTR NIDPFTSHLVEIAHMAEQDSVFEQSAHLHDLALPAKLAESTPDINGSRSL DIAGNRVRNIVERSEERFLRDHSEHAGSGFSDLELMTITADDVGRSVPEPLRG LGLSVRA"  
 1375..1398  
 /locus\_tag="Rv0282"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 2115..3731  
 /locus\_tag="Rv0283"  
 2115..3731  
 /locus\_tag="Rv0283"  
 /function="UNKNOWN"  
 /note="Rv0283, (MTV035.11), len: 538 aa. Possible conserved membrane protein, similar to several hypothetical mycobacterial proteins e.g. Z94121|MTV15F10.16|Rv3895c from Mycobacterium tuberculosis (495 aa), FASTA scores: opt: 698, E(): 0. (37.6% identity in 492 aa overlap); Rv1782; Rv3450c; Rv3869; and Y14967|MLCB628.16|MLCB628.17c from Mycobacterium leprae (481 aa), FASTA scores: opt: 672, E(): 1.5e-31, (37.2% identity in 506 aa overlap). Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental  
 misc\_feature  
 gene  
 CDS  
 /transl\_table=11  
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"  
 /protein\_id="CAAL17358.1"  
 /db\_xref="GI:2909472"  
 /translation="MTNQCHDHDHRRSFASRTPVANNPKVYVRGFVTRHQVTG WFWVNRITAGIALHDTMLVDLPTQSRVLMGLVITGLIGSFVSLRPNQAG SVANLADSTRATYVVRGSQLPLNLTSARLIVGRVSPPTVKSTELDQPRGNLIG ICAPERMVQNTSTANWTCVGLNAPSRGADGVGVTVIAGLEDTGARAALGPQO AVLDGAGTLLNDGKSPIDLADHANTSGLGLADVPAPRIASGIFNAIPAPPL TAPIIDGNPASFGVPAPIGAVSSYALKDSGKTSITVQYVAVLPGLQISFVLA AILNNNSYGLQPPRLGADVEAKLPSRVLDTRRYPSEVSLVDVTRDPTVCATWSK PYGATSSLTLAGSALPVDPAHTVELVGAGNGGVAIVRVALAAGTGYTFTVGGGPD APGASLFWISDTGVRYGIDNEPQGVAGKAGVAGLGNPPEVPIWNSVLISLVEGPT LSRADALLAHDTLVPSRPARPVSAEGYR"  
 3024..3047  
 /locus\_tag="Rv0283"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 3728..7720  
 /locus\_tag="Rv0284"  
 3728..7720  
 /locus\_tag="Rv0284"  
 /function="UNKNOWN"  
 /note="Rv0284, (MTV035.12), len: 1330 aa. Possible conserved membrane protein, similar to products of two adjacent Mycobacterium leprae genes, MLCB628.16c (744 aa) and MLCB628.15c (597 aa); and throughout its length to several large Mycobacterium tuberculosis proteins: Y3447c, Rv3870, Rv1784, etc. Y14967|MLCB628.15 (744 aa), FASTA scores: opt: 942, E(): 0. (33.8% identity in 730 aa overlap); Y14967|MLCB628.14 (597 aa), FASTA scores: opt: 613, E(): 3.1e-30, (31.7% identity in 615 aa overlap); Z94121|MTV15F10.17 (1236 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap); Z95389|MTCY77.19 (1236 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap). Contains three PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"  
 /protein\_id="CAAL17359.1"  
 /db\_xref="GI:2909473"  
 /translation="NSRLIFEARRELAPSSHQGTIIIEAPPELPRVIPPILLRRLAL YLIGLIVGMIVALVATGMVVISPTLFFPFFVILLAAATLYRGNDKMRTEVDAERA DYRLYSVRDNRIRAAQAEQASALWSPDPTALASVPSRRQMRDPDPPVLRV GHTYLTTLRVNDTADIEDLPVSHSALSLDTSQISGIVFTGIDLTQSPITVL GERQAVRLAWIAQAVTWDPTVLGVLAARDLEGRDMNKLWLPVDIPGRIDL GPARNLSDPDELIALLGVLDRAPTQPTDALRHLLI VDDPDYDLGASPLAVGR AGTVVHSCASAPHREQYSDPEKPLRVAGHAERWGTGQWPIADAQFSADEAAH LARRLSNDSPHTAGLSAATRGASTLLGIEDASLDVPAWAPRREREVRPI GTVGTGEPLMDLKDSEGGGPGHLMQMTGSKSQSLMSILLSTLTHSABLIVI YADFAGEAGSFRDFPQVAVISNAEKSLADFATLGRVEARREMLRAGKRV QGSANVLEVENIAAGHSIPPTLFWADEFTMLADHPEYAEFLDYVARGRSF RIHLFASQTLDVGKIDKNYATYRIGLKVASPSVRSQIIGVEDAHIESGKHGV GFLVPAGATPIRFSTVVDGYIEPTQAKAVVSPPEKLYFTAAVPEPQTVIAD PLEQDPRPKRLIATIGEOLARYGPAPQLWLPDLDETILPSAALAGVGPQWRW PIQEDTRPPEMRDPLVDARSAGNVIHGSPKSKSTALQTLFSLASLHSHSHEVS FYCLDYGQQLRALQDLAHVGSVASALEPERIRTFGELEQLLSRQOREYFRORGAN GSTDDGFEVFLVDNLYGFRDNTQFNTRNELLARVTELVNVLGYHGIHTTTP SWLEPLAMRDGLRLRLHLDADSVNRVVGALRRPADVPDHPQGRGUTMAEHF LFAPELDAQTNPVAAINARYPGMAAPRLLPTNLAPHAVGELYRFDQLVIGORE DIAVPLDANPLMVFCDGSGKTLRLHRIIRTEHSTADRVAFVDTLRLHLVD EPLFPDNEYTNIDRIIPAMGLGANIEARRPPAGMSAELSRLWTFAGHTYLIIDV DQVPSAMTGPYIGREWTPILGILAGDGLRVI VTGRATGSALHMTSLPLRRF NDLQATLMLAGNPADSKIRGERFARLPAGRAILLTDSFTYVQLINPLVDAVVS GTQKNGSQS"  
 5162..5185  
 /locus\_tag="Rv0284"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 6212..6235  
 /locus\_tag="Rv0284"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 7046..7069  
 /locus\_tag="Rv0284"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"



\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 222049: contig of 222049 bp in length
* 222050 222149: gap of unknown length
* 222150 236912: contig of 14763 bp in length
* 236913 237012: gap of unknown length
* 237013 301105: contig of 64093 bp in length
* 301106 301205: gap of unknown length
* 301206 305403: contig of 4198 bp in length
* 305404 305503: gap of unknown length
* 305504 352310: contig of 46807 bp in length
* 352311 352410: gap of unknown length
* 352411 353860: contig of 1450 bp in length
* 353861 353960: gap of unknown length
* 353961 355587: contig of 1627 bp in length
* 355588 355687: gap of unknown length
* 355688 357943: contig of 2256 bp in length
* 357944 358043: gap of unknown length
* 358044 360097: contig of 2053 bp in length
* 360097 360196: gap of unknown length
* 360197 361828: contig of 1632 bp in length
* 361829 361928: gap of unknown length
* 361929 364193: contig of 2365 bp in length
* 364194 364293: gap of unknown length
* 364294 374204: contig of 9911 bp in length.
  
```

## FEATURES

source

```

1. 374204
  /organism="Rattus norvegicus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10116"
  /clone="CH230-312K14"
  107683..110071
    /note="wgs_contig"
  220130..222049
    /note="wgs_contig"
  301206..303668
    /note="wgs_contig"
  303719..305403
    /note="wgs_contig"
  
```

## ORIGIN

```

Query Match      82.5%; Score 19.8; DB 2; Length 110000;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
  
```

```

QY 1 GGAGCGTTGGTACGTCGGAT 23
  
```

```

Db 98648 GGGGGCCTGGTACGTCGGAT 98626
  
```

## RESULT 8

AC142436

```

DEFINITION      AC142436 183433 bp DNA linear HTG 01-APR-2003
                  Rattus norvegicus clone CH230-77C16, *** SEQUENCING IN PROGRESS
                  ***, 61 unordered pieces.
  
```

ACCESSION AC142436

VERSION AC142436.2 GI:29423822

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
  
```

REFERENCE 1 (bases 1 to 183433)

```

AUTHORS      Murzyn D, Marie, Metzker M, Lee, Abramson S, Adams C, Alder J,
              Allen C, Allen H, Alsbrooks S, Amin A, Anguiano D,
              Anyalebechi V, Aoyagi A, Ayodeji M, Baca E, Baden H,
              Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F,
              Biswal K, Blair J, Blankenburg K, Blyth P, Brown M,
  
```

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegged, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Fu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaiz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlaczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, X., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 183433)

Worley, K.C.

Direct Submission

Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 183433)

Worley, K.C.

Direct Submission

Submitted (01-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Apr 1, 2003 this sequence version replaced gi:29366867.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: KEOJ

Center clone name: CH230-77C16

----- Summary Statistics

Sequencing vector: plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 161009 bases at least Q40

Consensus quality: 167462 bases at least Q30

Consensus quality: 172906 bases at least Q20





Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Housi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, J., Johnson, R., Jolivet, S., Joudah, S., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M. P., Meadow, M., Mei, G., Mezker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vexa, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 216124)  
Worley, K.C.

Direct Submission  
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 216124)  
Worley, K.C.

Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: f0J2  
Center clone name: CH230-1N4  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 132680 bases at least Q40  
Consensus quality: 138597 bases at least Q30  
Consensus quality: 142739 bases at least Q20  
Estimated insert size: 145045; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1  
67820: contig of 67820 bp in length  
67821: gap of unknown length  
71048: contig of 3128 bp in length  
71149: gap of unknown length  
130978: contig of 59830 bp in length  
130979: gap of unknown length  
131079: contig of 68392 bp in length  
199471: gap of unknown length  
199472: contig of 5646 bp in length  
205217: gap of unknown length  
205218: contig of 1532 bp in length  
206848: gap of unknown length  
206849: contig of 1470 bp in length  
208419: gap of unknown length  
208420: contig of 2113 bp in length  
210632: gap of unknown length  
210732: contig of 3370 bp in length  
214101: gap of unknown length  
214102: contig of 1923 bp in length.  
214202: Location/Qualifiers  
1. 216124

#### FEATURES

source

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10115"  
/clone="CH230-1N4"

#### ORIGIN

Query Match 82.5%; Score 19.8; DB 2; Length 216124;  
Best Local Similarity 91.3%; Pred. No. 1.3e-02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGGGCTTGGTACGTCGCAT 23

Db 183360 GGGGGCTTGGTACGTCGCAT 183382

#### RESULT 10

AC129229

LOCUS

DEFINITION

AC129229

AC129229

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayala, M., Banks, T.,

Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burckett, C., Burrell, K.L., Byrd, N.C.,

Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Deuchwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, K., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

AC129229 276158 bp DNA linear HTG 09-OCT-2002  
Rattus norvegicus clone CH230-1L6, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
6 unordered pieces.

AC129229

AC129229.2 GI:23195757

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 276158)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayala, M., Banks, T.,

Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burckett, C., Burrell, K.L., Byrd, N.C.,

Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Deuchwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, K., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,



Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K., Jolivet,S., Joudah,S., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherz,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonake,P., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 276158)

Worley,K.C.

Direct Submission

Submitted (28-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 276158)

Worley,K.C.

Direct Submission

Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:22000833. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/at/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence contigs will be indicated in the feature table.

----- Genom Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: TUGA

Center clone name: CH230-1L6

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 247882 bases at least Q40

Consensus quality: 252477 bases at least Q30

Consensus quality: 254863 bases at least Q20

Estimated insert size: 303705; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will

\* be preserved.

\* 1 255814: contig of 255814 bp in length  
\* 255815 255914: gap of unknown length  
\* 255915 269133: contig of 13599 bp in length  
\* 269134 269514: gap of unknown length  
\* 269515 270984: contig of 1371 bp in length  
\* 270985 271084: gap of unknown length  
\* 271085 272870: contig of 1786 bp in length  
\* 272871 272970: gap of unknown length  
\* 272971 274397: contig of 1827 bp in length  
\* 274398 274897: gap of unknown length  
\* 274898 276158: contig of 1261 bp in length.

## FEATURES

## source

Location/Qualifiers

1..276158

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-1L6"

1..2090

/note="wgs\_contig"

2141..4680

/note="wgs\_contig"

264601..269513

/note="wgs\_contig"

## ORIGIN

Query Match 82.5%; Score 19.8; DB 2; Length 276158;

Best Local Similarity 91.3%; Pred. No. 1.3e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGGCGTTGGGTACGTCGCAT 23

Db 226654 GGGGGCTTTGGGTACGTCGCAT 226676

## RESULT 11

## NCB8G12/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"

## 1..154038

## /organism="Neurospora crassa"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:5141"



```

WMYKSPKAITAIYKYKTRLDRSKYVLDCLMDVMQVNSDFEKPPTPDPCKNNHHHHT
NHNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
KKRGSDESKVLG"
/complement(21351..22913)
/gene="B8G12.060"
/number=1
repeat_region
21391..21423
/note="33 bp gttggtggtggtggtggtg tandem repeat"
repeat_region
21446..21469
/note="24 bp tgttgttgtgt tandem repeat"
repeat_region
21506..21538
/note="33 bp tgttgttgtgttgttgtgt tandem repeat"
gene
24030..24701
/gene="B8G12.070"
gene
24030..24701
/gene="B8G12.070"
complement(24030..24701)
/gene="B8G12.070"
CDS
complement(24030..24701)
/gene="B8G12.070"

Query Match      80.08;   Score 19.2; DB 8; Length 154038;
Best Local Similarity    87.58; Pred.No.2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GGAGCGGTGGGTACGTCGCATC 24
|||||
Db      84962 GGAGCGGTGGGTATGCTAGCGTC 84939

RESULT 12
MLEPRTN9/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Cole S.T., Eigleier K., Parkhill J., James K.D., Thomson N.R., Wheeler P.R., Honore N., Ganier T., Churcher C., Harris D., Mungall K., Basham D., Brown J.D., Chillingworth T., Connor R., Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N., Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S., Murphy K.M., Oliver Quail M.A., Rajadream M.-A., Rutherford K., Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Taylor K., Whitthead S., Woodward J.R. and Barrrell B.G.
Massive gene decay in the leprosy bacillus
Nature 409 (6823), 1007-1011 (2001)
21128732
PUBMED
11234002
REFERENCE
2 (bases 1 to 308950)
Parkhill,J.
Direct Submission
Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75794, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
Notes:
Details of M. leprae sequencing at the Sanger Centre are available from http://www.sanger.ac.uk/projects/M_leprae/A_relational_database_containing_the_M._leprae_sequences_is_available_from_http://genolist.pasteur.fr/leproma/
Location/Qualifiers
1..308950
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/strain="TN"
/db_xref="taxon:1769"
86..1691
gene

```

```

misc_feature
AVADWAMFNWPIKTPGLRCQTDVAPQNNRLPAYRSIAAPVLVIGFAEDVVTTPSL
GREVADVLNPGRYLQIPDAGHLGFFERPEAVNAALQFFAGVKGEW"
2784. 3434
/ gene="ML2269"
/ note="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 86.00, E-value 7.7e-22"
3705. 5375
/ gene="mend"
/ note="synonym: ML2270"
3705. 5375
/ gene="mend"
/ note="Similar to Mycobacterium tuberculosis menD or
RV0555 or MTCY25D10.34 TR:006421 (EMBL:Z95558) (554 aa)
fasta scores: E(): 0, 85.4% id in 556 aa. Similar to
Bacillus subtilis menaquinone biosynthesis protein Mend
(includes 2-succinyl-6-hydroxy-
2,4-cyclohexadiene-1-carboxylate synthase, 2-oxoglutarate
decarboxylase) mend SW:MEND_BACSU (P23970; P23969; O34492)
(580 aa) fasta scores: E(): 2.9e-21, 31.6% id in 594 aa."
/ codon_start=1
/ trans1_table=11
/ product="putative
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate
synthase / 2-oxoglutarate decarboxylase"
/ protein_id="CAC31786.1"
/ db_xref="GI:13093908"
/ db_xref="GOA:Q9CBB0"
/ db_xref="SPTREMBL:Q9CBB0"
/ translation="MNPSTQAVVVDLIRGGVDRVVLCPGSRNAPLAFADADHA
GRILHVIDERTAGYLAIGLAAAGAPCVAMTSGTAVANLGPVAVNAYRVLIV
LSANRPVELLGTGANTHEQLGYEGTQVRAISLGLAEDAPERLDSLNATWRSATRV
LVAATGRTANAGVQDIPUREPLIDPEPHGAIPTQGRPGKPWTYIPQVTFDQPL
EIDSATPVIAGHAGVHNPETLPIVAEPTAFPGYRPLPLPLALPLLPKQVIM
LGRFTLRVTSALLADPQVFPYALTTCPCWFDVSGNSQAAGTAVTTGTPHAWLQRC
AEMNQHAISTRVSOLAAHPILITGLHVAADALRPGDQVLGASNPVDMALVGLST
GLIOVRNKGAGIDGTVTAIGAAVAYSPGVAGTDRPARTVALIGLAFVHDSS
GLLIGPTEPTQOLKIVVNDNGGIFELLEQDPRLSVSSRIEFGTHPDVDVGLCR
AYHVEGQIEVDKLEALDEPGVGRVLEVAADRSSLRQLHRAITAIL"
5386. 5392
/ note="possible RBS"
5403. 5930
/ gene="ML2271"
5403. 5930
/ gene="ML2271"
/ note="ML2271"
/ note="Similar to Mycobacterium tuberculosis putative
membrane protein RV0556 or MTCY25D10.35 TR:006422
(EMBL:Z95558) (171 aa) fasta scores: E(): 0, 79.2% id in
168 aa. Contains hydrophobic, possible membrane-spanning
regions."
/ codon_start=1
/ trans1_table=11
/ product="putative membrane protein."
/ protein_id="CAC31787.1"
/ db_xref="GI:13093909"
/ db_xref="SPTREMBL:Q9CBA9"
/ translation="MILSKILLRILLHGRDEPSTRVIMRWGRIAVLIVTGLITL
QSVLLVAGANNDITIQHNMGVAQAEVLSPRRSTIEFTVPTVRPELVYLSE
LSTGMRYVVENKNDPLNVRVHRNAGLAIIPAGTAVVCWLATVVLVALVLDKRL
DRHTESAVRSQPTS"
6036. 7128
/ gene="ML2272"
/ note="Possible pseudogene of M. tuberculosis orthologue
RV0557 (Best blastx score 351)"
/ pseudo
/ codon_start=1
/ trans1_table=11
/ product="conserved hypothetical protein (pseudogene)"
7183. 7899
/ gene="ML2273"
7183. 7899

```

```

/ gene="ML2273"
/ note="Similar to Mycobacterium tuberculosis putative
ubiquinone/menaquinone biosynthesis methyltransferase
RV0558 or MTCY25D10.37 TR:006424 (EMBL:Z95558) (234 aa)
fasta scores: E(): 0, 83.3% id in 228 aa. Similar to many
e.g. Streptomyces coelicolor putative
ubiquinone/menaquinone methyltransferase SC10A7.27C
TR:Q9XAP6 (EMBL:AL078618) (231 aa) fasta scores: E(): 0,
64.9% id in 225 aa and to Escherichia coli
ubiquinone/menaquinone biosynthesis methyltransferase UbiE
ubE SW:UBIE_ECOLI (P27851) (251 aa) fasta scores: E():
80.0%; Score 19.2; DB 1; Length 308950;
Query Match
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGAGGCGTTCGGTACGTCGCATC 24
Db 238182 GGAGGCGTTCGGTACTGTGCGTC 238159

RESULT 13
AC079436 174622 bp DNA linear ROD 01-FEB-2003
Rattus norvegicus chromosome 4 clone RP31-108L19 strain Brown
Norway, complete sequence.
AC079436
AC079436.2 GI:17530715
HTG.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 174622)
AUTHORS Akhter N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Graniter, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 174622)
Direct Submission
Submitted (01-SEP-2000) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 174622)
Green, E.D.
Direct Submission
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
4 (bases 1 to 174622)
Green, E.D.
Direct Submission
Submitted (01-FEB-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Dec 12, 2001 this sequence version replaced gi:9958048.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@hgr.nih.gov
----- Project Information
Center project name: rd
Center clone name: 108L19
This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an

```

alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features section.

#### FEATURES

source  
Location/Qualifiers  
1. .174622  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="Brown Norway"  
/db\_xref="taxon:10116"  
/chromosome="4"  
/clone="RP31-108L19"  
/clone\_lib="RP31"  
54910. .54979  
/note="single clone coverage"  
88104. .88667  
/note="single clone coverage"  
156421. .156425  
/note="low quality single stranded/single chemistry region"  
156436. .156440  
/note="low quality single stranded/single chemistry region"  
165422. .165426  
/note="low quality single stranded/single chemistry region"  
165476  
/note="low quality single stranded/single chemistry region"  
165479. .165480  
/note="low quality single stranded/single chemistry region"  
165495. .165501  
/note="low quality single stranded/single chemistry region"  
165677. .165679  
/note="low quality single stranded/single chemistry region"  
165688. .165693  
/note="low quality single stranded/single chemistry region"  
165708. .165713  
/note="low quality single stranded/single chemistry region"  
165742. .165746  
/note="low quality single stranded/single chemistry region"  
165768. .165769  
/note="low quality single stranded/single chemistry region"  
165781. .165784  
/note="low quality single stranded/single chemistry region"  
165940. .174622  
/note="clone overlaps with GenBank Accession Number AC079990 (nucleotides 1-8683) clone RP31-327J16 (center project name rc); this annotated segment represents overlap with the finished portion of RP31-327J16 presented in AC079990; overlap with the unfinished portion of RP31-327J16 extends to nucleotide 148342"

#### ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 174622;  
Best Local Similarity 90.9%; Pred. No. 3.9e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGCGGTGGGTACGGTCGCA 22  
||||| ||||||| ||||||| |||||||  
DB 81951 GGAGGCCCTGGGTACGGTCGCA 81930

#### RESULT 14

AC106493/c  
LOCUS AC106493 232195 bp DNA linear HTG 21-SEP-2002  
DEFINITION Rattus norvegicus clone C230-183H3, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 7 unordered pieces.  
AC106493  
VERSION AC106493.3 GI:23265214  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

#### REFERENCE

1 (bases 1 to 232195)  
Muzny, D. Mazie., Metsker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geet, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorushtewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Manghney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munica, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajds, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinkle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmari, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

#### TITLE

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 232195)  
AUTHORS Worley, K. C.  
TITLE Direct Submission



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 1992.42 Seconds

Sequence: 1 ggaagcggttggttaagtcgcac 24  
(without alignments)  
378.717 Million cell updates/sec

Title: US-10-624-714-7

Perfect score: 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_esti:\*

9: gb\_est2:\*

10: gb\_est3:\*

11: gb\_est4:\*

12: gb\_est5:\*

13: gb\_est6:\*

14: gb\_est7:\*

15: gb\_est8:\*

16: gb\_est9:\*

17: gb\_est10:\*

18: gb\_est11:\*

19: gb\_est12:\*

20: gb\_est13:\*

21: gb\_est14:\*

22: gb\_est15:\*

23: gb\_est16:\*

24: gb\_est17:\*

25: gb\_est18:\*

26: gb\_est19:\*

27: gb\_est20:\*

28: gb\_est21:\*

29: gb\_est22:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	19.8	82.5	590	28	BH268649 CH230-186
c 2	19.8	82.5	625	28	BH288188 CH230-77C
c 3	19.2	80.0	594	12	BM638934
4	19.2	80.0	647	28	BH382141 AG-ND-124

5	19.2	80.0	661	12	BM635844
6	19.2	80.0	993	29	CNS01J7Y
7	19.2	80.0	1023	29	CNS01QL5
c 8	18.4	76.7	725	29	AG135453
9	18.2	75.8	384	10	BE946157
10	18.2	75.8	452	28	AQ443923
c 11	18.2	75.8	521	12	EG827233
c 12	18.2	75.8	641	10	BB664588
c 13	18.2	75.8	668	28	BH341450
14	18.2	75.8	683	28	AZ512348
15	18.2	75.8	784	13	BQ85217
16	17.8	74.2	252	12	EM068683
17	17.8	74.2	589	14	CA235402
c 18	17.8	74.2	708	28	BH948951
c 19	17.8	74.2	718	13	BQ802895
c 20	17.8	74.2	816	28	AQ858532
c 21	17.8	74.2	1119	12	BG107350
c 22	17.6	73.3	188	10	BBS75037
c 23	17.6	73.3	227	10	BBS71026
c 24	17.6	73.3	228	13	BY173995
c 25	17.6	73.3	254	10	BBS38158
c 26	17.6	73.3	260	9	AL909710
c 27	17.6	73.3	287	10	BBS721896
c 28	17.6	73.3	288	9	AV050731
c 29	17.6	73.3	290	9	AV050731
c 30	17.6	73.3	341	12	BI202226
c 31	17.6	73.3	369	10	BBS81654
c 32	17.6	73.3	390	13	BY036704
c 33	17.6	73.3	392	14	CF486973
34	17.6	73.3	412	9	AJ283991
c 35	17.6	73.3	472	29	CE503180
c 36	17.6	73.3	482	10	BBS82823
c 37	17.6	73.3	495	29	CG032539
c 38	17.6	73.3	496	10	BBS83326
c 39	17.6	73.3	519	14	CA184364
c 40	17.6	73.3	551	13	BQ591994
c 41	17.6	73.3	575	13	BQ061589
c 42	17.6	73.3	600	13	BU921446
c 43	17.6	73.3	615	14	CA168966
44	17.6	73.3	625	13	CA109750
c 45	17.6	73.3	656	12	BM588137

## ALIGNMENTS

RESULT 1  
BH268649/c  
LOCUS  
DEFINITION  
CH230-186D22-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-186D22, genomic survey sequence.  
ACCESSION  
BH268649  
VERSION  
BH268649.1 GI:17180955  
KEYWORDS  
GSS.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 590)  
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,  
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,  
Riggs,F., de Jong,P. and Fraser,C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)  
Other GSSs: CH230-186D22.TJ  
Contact: Shaving Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org



Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).  
Clones may be purchased from BACPAC Resources  
(<http://www.chori.org/bacpac/oreringinformation.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)  
Plate: 186 row: D column: 22  
Seq primer: I7  
Class: BAC ends.

## FEATURES

source  
1. .590  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SENHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-186D22"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
Pieter de Jong"

## ORIGIN

Query Match 82.5%; Score 19.8; DB 28; Length 590;  
Best Local Similarity 91.3%; Pred. No. 9.7e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGAGCGCTTGGGTACGTCGCAT 23

Db 357 GGGGGCCTTGGGTACGTCGCAT 335

## RESULT 2

BH288188/c  
LOCUS  
DEFINITION CH230-77C16.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-77C16, genomic survey sequence.

ACCESSION BH288188

VERSION BH288188.1 GI:17200596

KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)

ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 625)

AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsagaye,G., Geer,K.,  
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,  
Riggs,F., de Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment

JOURNAL Unpublished (1999)

COMMENT Other GSSs: CH230-77C16.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/oreringinformation.htm>). BAC end

page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)

Plate: 77 row: C Column: 16

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .625

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/strain="BN/SENHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-77C16"  
/sex="Female"

/cell\_type="Brain"

/clone\_lib="CHORI-230 Segment 1"

/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by

Pieter de Jong"

## ORIGIN

Query Match 82.5%; Score 19.8; DB 28; Length 625;  
Best Local Similarity 91.3%; Pred. No. 9.8e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGAGCGCTTGGGTACGTCGCAT 23

Db 374 GGGGGCCTTGGGTACGTCGCAT 352

## RESULT 3

BH288188/c

LOCUS

DEFINITION

17000687567070 A.Gam.ad.cdNAL Anopheles gambiae cDNA clone

19600449621708 5', mRNA sequence.

ACCESSION BH288188

VERSION BH288188.1 GI:18938445

KEYWORDS EST.

SOURCE

ORGANISM

Anopheles gambiae (African malaria mosquito)

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

REFERENCE 1 (bases 1 to 594)

AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,

Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.

Celera Anopheles gambiae EST project

Unpublished (2002)

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: [HoltRA@celera.com](mailto:HoltRA@celera.com)

Plate: NU01004N8J row: K column: 18

Seq primer: M13 Reverse.

Location/Qualifiers

1. .594

/organism="Anopheles gambiae"

/mol\_type="mRNA"

/strain="RSP-ST (Reduced susc. to Permethrin - std.

chromosome)"

/db\_xref="taxon:7165"

/clone="19600449621708"

/dev\_stage="Adult"

/lab\_host="DH10b"

/clone\_lib="A.Gam.ad.cdNAL"

/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole

adult mosquitoes (mixed sex) frozen on liquid nitrogen.

cDNA inserts >500 bp cloned directionally into pSport 1.

Not 1 site is 3'. Clones available through the Malaria

Research and Reference Reagent Resource Center

([www.malaria.mr4.org](http://www.malaria.mr4.org)).

Query Match 80.0%; Score 19.2; DB 12; Length 594;

Best Local Similarity 87.5%; Pred. No. 1.7e+03;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGCGCTTGGGTACGTCGCATC 24

Db 490 GGAGGTGTTGAGTACGGCCGCATC 513



```

RESULT 4
BH382141
LOCUS
DEFINITION
AG-ND-124C16.TF ND-TAM Anopheles gambiae genomic clone
AG-ND-124C16, genomic survey sequence.
ACCESSION
BH382141
VERSION
BH382141.1 GI:17328283
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 647)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,B.R., Carille,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (5), 720-728 (2003)
JOURNAL
MEDLINE
22542063
PUBMED
12553398
COMMENT
Other GSSs: AG-ND-124C16.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@igr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..647
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-124C16"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

```

# FEATURES

```

source
Query Match 80.0%; Score 19.2; DB 28; Length 647;
Best Local Similarity 87.5%; Pred. No. 1.7e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

# ORIGIN

```

Qy 1 GGAGGCGTTGGTACGTCGCATC 24
|||||
Db 202 GGAGGTGTGAGTACGCGGCATC 225
|||||

```

```

RESULT 5
BM635844
LOCUS
DEFINITION
17000687560517 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
19600449642832 5', mRNA sequence.
ACCESSION
BM635844
VERSION
BM635844.1 GI:18935355
KEYWORDS
SOURCE
EST.
Anopheles gambiae (African malaria mosquito)
Anopheles gambiae

```

# REFERENCE

```

AUTHORS
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01003CY3 row: K column: 22
Seq primer: M13 Reverse.
Location/Qualifiers
1..661
/organism="Anopheles gambiae"
/mol_type="cDNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449642832"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cdna1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

```

# FEATURES

```

source
Query Match 80.0%; Score 19.2; DB 12; Length 661;
Best Local Similarity 87.5%; Pred. No. 1.7e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

# ORIGIN

```

Qy 1 GGAGGCGTTGGTACGTCGCATC 24
|||||
Db 506 GGAGGTGTGAGTACGCGGCATC 529
|||||

```

# RESULT 6

```

CNS01J7Y
LOCUS
DEFINITION
Anopheles gambiae GSS SP6 end of clone 12N08 of Notredame1 library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION
AL146703
VERSION
AL146703.1 GI:7004849
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 993)
Genoscope.
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 993)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Direct Submission
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

# REFERENCE

```

AUTHORS
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

# REFERENCE

```

AUTHORS
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

```

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
1 (bases 1 to 661)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,  
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished (2002)  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01003CY3 row: K column: 22  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..661  
/organism="Anopheles gambiae"  
/mol\_type="cDNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449642832"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
1 (bases 1 to 661)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,  
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished (2002)  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01003CY3 row: K column: 22  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..661  
/organism="Anopheles gambiae"  
/mol\_type="cDNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449642832"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
1 (bases 1 to 661)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,  
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished (2002)  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01003CY3 row: K column: 22  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..661  
/organism="Anopheles gambiae"  
/mol\_type="cDNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449642832"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
1 (bases 1 to 661)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,  
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished (2002)  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01003CY3 row: K column: 22  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..661  
/organism="Anopheles gambiae"  
/mol\_type="cDNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449642832"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

```

FEATURES
  source
    Location/Qualifiers
      1..993
      /organism="Anopheles gambiae"
      /mol_type="genomic DNA"
      /strain="PEST"
      /db_xref="taxon:7165"
      /clone="12N08"
      /clone_lib="NotreDamel"
      /note="end : SP6"

ORIGIN
  Query Match      80.0%; Score 19.2; DB 29; Length 993;
  Best Local Similarity 87.5%; Pred. No. 1.8e+03;
  Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGTCGCATC 24
    |||||
DB 148 GGAGGTGTTGAGTACGCGCCGATC 171
    |||||

RESULT 7
CNS01Q15 1023 bp DNA linear GSS 14-JUN-2001
LOCUS Anopheles gambiae GSS SP6 end of clone 31B11 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL156249
VERSION AL156249.1 GI:7017168
KEYWORDS GSS.
ORGANISM Anopheles gambiae (African malaria mosquito)
SOURCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 1023)
  AUTHORS Direct Submission
  TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
  JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  2 (bases 1 to 1023)
  Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  Direct Submission
  TITLE Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
  JOURNAL Roux, Paris 75015, France
  COMMENT This clone is from an A. gambiae BAC library provided by F.H.
  Collins and sequenced by Genoscope in collaboration with the
  Laboratory of Biochem. and Biol. Molec. of Insects, Institut
  Pasteur.

FEATURES
  source
    Location/Qualifiers
      1..1023
      /organism="Anopheles gambiae"
      /mol_type="genomic DNA"
      /strain="PEST"
      /db_xref="taxon:7165"
      /clone="31B11"
      /clone_lib="NotreDamel"
      /note="end : SP6"

ORIGIN
  Query Match      80.0%; Score 19.2; DB 29; Length 1023;
  Best Local Similarity 87.5%; Pred. No. 1.8e+03;
  Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGTCGCATC 24
    |||||
DB 879 GGAGGTGTTGAGTACGCGCCGATC 902
    |||||

RESULT 8
AG135453/c 725 bp DNA linear GSS 04-NOV-2001
LOCUS Anopheles gambiae GSS clone: PTB-148L19.F, genomic survey sequence.
DEFINITION Pan troglodytes DNA, clone: PTB-148L19.F, genomic survey sequence.

```

```

ACCESSION AG135453
VERSION AG135453.1 GI:16665131
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
  AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
  Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE BAC end sequences of Library PTB
  JOURNAL Unpublished
  REFERENCE 2 (bases 1 to 725)
  AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
  Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE Direct Submission
  JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
  (E-mail:chimbesc@sc.riken.go.jp URL:http://hgp.gsc.riken.go.jp/,
  Tel:81-45-503-9111, Fax:81-45-503-9170)
  COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
  was generated during the R&D process and may have higher chance of
  clone tracking errors.
  PRIMERS
  Sequencing: -21M13
  LIBRARY
    Vector : pKS145
    R.Site 1 : SacI
    R.Site 2 : SacI.
  FEATURES
    source
      Location/Qualifiers
        1..725
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-148L19.F"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      76.7%; Score 18.4; DB 29; Length 725;
  Best Local Similarity 95.0%; Pred. No. 3.6e+03;
  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGTCG 20
    |||||
DB 207 GGAGCGTTGGTACGCG 188
    |||||

RESULT 9
BE946157 384 bp mRNA linear EST 03-OCT-2000
LOCUS UI-M-B20-axs-c-02-0-UI-s1 NIH BMAP.MH12 Mus musculus cDNA clone
DEFINITION UI-M-B20-axs-c-02-0-UI 3', mRNA sequence.
ACCESSION BE946157
VERSION BE946157.1 GI:10523881
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 384)
  AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE Normalization and subtraction: two approaches to facilitate gene
  JOURNAL discovery
  MEDLINE Genome Res. 6 (9), 791-806 (1996)
  PUBMED 9704477
  COMMENT 889548
  Contact: Chin, H
  National Institute of Mental Health
  6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
  20892-9643, USA

```

Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the hippocampus tissue cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BVAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BVAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

#### FEATURES source

Location/Qualifiers  
1..384  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-B20-axs-c-02-0-UI"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BVAP MH12"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BVAP MH12 library is derived from mouse hippocampus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.  
TAG TISSUE=hippocampus  
TAG\_LIB=NIH\_BVAP\_MH12  
TAG\_SEQ=CGGTA"

#### ORIGIN

Query Match 75.8%; Score 18.2; DB 10; Length 384;  
Best Local Similarity 87.0%; Pred. No. 3.9e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTGCAT 23  
|||||  
DB 334 GGAGCGTTGGACGGTTCAT 356  
|||||

#### RESULT 10

AQ443923 452 bp DNA linear GSS 09-JAN-2001  
GSSTC01378 Trypanosoma cruzi random genomic library Trypanosoma cruzi random genomic clone G16H2, genomic survey sequence.

ACCESSION AQ443923.2 GI:9293389

VERSION GSS.

KEYWORDS Trypanosoma cruzi

SOURCE Trypanosoma cruzi

ORGANISM Trypanosoma cruzi

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

1 (bases 1 to 452)

Agarero, F., Verdun, R., Frasch, A.C.C. and Sanchez, D.O.

A random sequencing approach for the analysis of the trypanosoma

cruzi genome: General structure, large gene and repetitive DNA

families, and gene discovery

Genome Res. 10 (12), 1996-2005 (2000)

20568489

11116094

On Jul 19, 2000 this sequence version replaced gi:4555387.

Contact: Sanchez D.O.

Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral

San Martin)

Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos

Aires, Argentina

Tel: (54-11) 4580/7255/7

Fax: (54-11) 4752-9639  
Email: csanchez@lib.unsam.edu.ar  
Seq primer: T7  
Class: shotgun.

#### FEATURES

##### source

Location/Qualifiers  
1..452  
/organism="Trypanosoma cruzi"  
/mol\_type="genomic DNA"  
/strain="CL-Srener"  
/db\_xref="taxon:5693"  
/clone="G16H2"  
/cell\_type="epimastigote"  
/clone\_lib="Trypanosoma cruzi random genomic library"  
/notes="Vector: PBS(-) (Stratagene); T. cruzi DNA was randomly sheared using a nebulizer and the 1 to 2 Kb range was gel purified and cloned into the dephosphorylated HincII site of the vector"

#### ORIGIN

Query Match 75.8%; Score 18.2; DB 28; Length 452;  
Best Local Similarity 87.0%; Pred. No. 4e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTGCAT 23  
|||||  
DB 344 GGAGCGTTGGTACGGAGCAT 366  
|||||

#### RESULT 11

BG827233/c

LOCUS

DEFINITION

602749369F1 NIH\_MGC\_17 Homo sapiens CDNA clone IMAGE:4902233 5',

mRNA sequence.

ACCESSION BG827233

VERSION BG827233.1 GI:14174820

KEYWORDS ESI.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 521)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-x@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI798 row: i column: 18

High quality sequence stop: 113.

Location/Qualifiers

1..521

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4902233"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_17"

/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;

Site 2: XhoI; CDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGACGAG(G). Size-selected >500bp

for average insert size 1.8kb library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

#### ORIGIN

```

/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate
lung"
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGCACTCGAGTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATAATTAATCCCGCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN
Query Match 75.8%; Score 18.2; DB 10; Length 641;
Best Local Similarity 87.0%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGGCGTTGGTACGGTGCAT 23
Db 596 GGAGGCGTTGGTACCAATCTCAT 574

RESULT 13
LOCUS BH341450 668 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-41H20-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
BH341450
VERSION BH341450.1 GI:17272184
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 668)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
OTHER GSSs: CH230-41H20.TJ
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.html). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 41 row: H column: 20
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..668
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-41H20"

Query Match 75.8%; Score 18.2; DB 12; Length 521;
Best Local Similarity 87.0%; Pred. No. 4.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGGCGTTGGTACGGTGCAT 23
Db 407 GGAGGCGTTGGTACTGCGCGT 385

RESULT 12
LOCUS BB664588/c 641 bp mRNA linear EST 26-OCT-2001
DEFINITION BB664588 RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030047D23 5', mRNA sequence.
ACCESSION BB664588
VERSION BB664588.1 GI:16498342
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 641)
AUTHORS Hara,A., Hiramoto,K., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arawaka,T., et al. 2001)
Unpublished (2001)
CONTACT: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1..641
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="E030047D23"
/tissue_type="lung"

```

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 75.8%; Score 18.2; DB 28; Length 668;  
Best Local Similarity 87.0%; Pred. No. 4.2e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGGCGTTGGGTACGGTCGCAT 23  
|||||  
DB 356 GGGGCGCTGGGTTGGTCGCAT 334  
|||||

RESULT 14  
AZ512348  
LOCUS  
DEFINITION  
1M0357M17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0357M17 R, genomic survey sequence.

ACCESSION  
AZ512348  
VERSION  
GSS.  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 683)

REFERENCE  
AUTHORS  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von  
Niederhausern, A. and Wright, D., Weiss, R.

TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL  
COMMENT  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0357 row: M column: 17  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 683.

## FEATURES

source

1. .683  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0357M17"  
/sex="Male"

/lab\_host="F2. Coli strain XL10-Gold, Tl-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number

## ORIGIN

Query Match 75.8%; Score 18.2; DB 28; Length 683;  
Best Local Similarity 87.0%; Pred. No. 4.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGGCGTTGGGTACGGTCGCATC 24  
|||||  
DB 475 GAGGCGTTGGGTACGGTCGCATC 497  
|||||

## RESULT 15

BQ885217  
LOCUS  
DEFINITION  
BQ885217 784 bp mRNA linear EST 16-AUG-2002  
AGENCOURT\_8750502 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6332872  
5', mRNA sequence.

ACCESSION  
BQ885217  
VERSION  
EST.  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 784)

REFERENCE  
AUTHORS  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-x@mail.nih.gov

Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
Ph.D.

cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLAM13791 row: C column: 17  
High quality sequence stop: 389.

## FEATURES

source

1. .784  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6332872"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_130"  
/note="Organ: oocytes; Vector: pCMV-SPORT6.1; Site: 1:  
EcoRV; Site: 2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size 1.95 kb. Constructed by  
ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 75.8%; Score 18.2; DB 13; Length 784;  
Best Local Similarity 87.0%; Pred. No. 4.3e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGGCGTTGGGTACGGTCGCATC 24  
|||||  
DB 279 GAGGCGTTGGGTACGGTCGCATC 301  
|||||

Search completed: April 29, 2004, 11:37:05  
Job time : 1896.42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 42.3529 Seconds  
(without alignments)  
314.472 Million cell updates/sec

Title: US-10-624-714-7

Perfect score: 24

Sequence: 1 ggaggcgttggtacggtcgatc 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	24	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
3	17.2	71.7	4695	2 US-08-231-153A-57	Sequence 57, Appl
4	17.2	71.7	4695	2 US-08-486-273A-57	Sequence 57, Appl
5	17.2	71.7	4695	3 US-08-940-086A-57	Sequence 57, Appl
6	17.2	71.7	4695	4 US-08-940-035A-57	Sequence 57, Appl
7	17.2	71.7	4695	4 US-08-935-105A-57	Sequence 57, Appl
8	17.2	71.7	4695	4 US-08-648-797-57	Sequence 57, Appl
9	17.2	71.7	4695	4 US-09-386-123-57	Sequence 57, Appl
10	17.2	71.7	1230025	4 US-09-198-452A-1	Sequence 1, Appli
11	16.6	69.2	68750	3 US-09-335-409-1	Sequence 1, Appli
12	16.6	69.2	68750	4 US-09-568-102-1	Sequence 1, Appli
13	16.6	69.2	68750	4 US-09-567-969-1	Sequence 1, Appli
14	16.6	69.2	68750	4 US-09-568-480-1	Sequence 1, Appli
15	16.6	69.2	68750	4 US-09-568-486-1	Sequence 1, Appli
16	16.6	69.2	68750	4 US-09-568-472-1	Sequence 1, Appli
17	16.6	69.2	68750	4 US-09-567-899-1	Sequence 1, Appli
18	16	66.7	744	4 US-09-483-039A-4080	Sequence 480, Ap
19	16	66.7	765	4 US-09-489-039A-4465	Sequence 4465, Ap
20	16	66.7	845	4 US-09-976-594-128	Sequence 128, App
21	16	66.7	954	3 US-09-418-641-3	Sequence 3, Appli
22	16	66.7	43280	2 US-08-804-227C-1	Sequence 1, Appli
23	16	66.7	4403765	3 US-09-103-840A-2	Sequence 2, Appli
24	16	66.7	4411529	3 US-09-103-840A-1	Sequence 1, Appli
25	15.6	65.0	528	4 US-09-252-991A-5077	Sequence 5077, Ap
26	15.6	65.0	999	5 PCT-US91-00899-10	Sequence 10, Appl
27	15.6	65.0	1136	1 US-08-395-800A-9	Sequence 9, Appli

C 28	15.6	65.0	1155	1 US-08-434-151-1	Sequence 1, Appli
C 29	15.6	65.0	1155	1 US-08-208-889A-1	Sequence 1, Appli
C 30	15.6	65.0	1155	2 US-08-433-271-1	Sequence 1, Appli
C 31	15.6	65.0	1155	2 US-08-715-269-1	Sequence 1, Appli
C 32	15.6	65.0	1174	5 PCT-US95-07554-3	Sequence 3, Appli
C 33	15.6	65.0	1199	1 US-08-395-800A-5	Sequence 5, Appli
C 34	15.6	65.0	1467	4 US-09-252-991A-5167	Sequence 5167, Ap
C 35	15.6	65.0	1536	4 US-09-252-991A-5048	Sequence 5048, Ap
C 36	15.6	65.0	1599	4 US-09-252-991A-5013	Sequence 5013, Ap
C 37	15.6	65.0	2268	3 US-08-675-773B-4	Sequence 4, Appli
C 38	15.6	65.0	3373	1 US-08-273-411-2	Sequence 2, Appli
C 39	15.6	65.0	3791	3 US-08-675-773B-3	Sequence 3, Appli
C 40	15.6	65.0	8174	1 US-07-914-281-5	Sequence 5, Appli
C 41	15.6	65.0	8174	1 US-08-393-246-5	Sequence 5, Appli
C 42	15.6	65.0	8174	1 US-08-525-058A-5	Sequence 5, Appli
C 43	15.6	65.0	8174	2 US-08-696-731-5	Sequence 5, Appli
C 44	15.6	65.0	8174	3 US-09-042-531-5	Sequence 5, Appli
C 45	15.6	65.0	8174	5 PCT-US91-00899-3	Sequence 3, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 100.0%; Score 24; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred No. 0.038;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGGTACGTCGATC 24

Db 567638 GGAGCGTTGGGTACGTCGATC 567661

##### RESULT 2

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 100.0%; Score 24; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.038;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAGCGCTGGTACGTCGCATC 24  
Db 566196 GGAGCGCTGGTACGTCGCATC 566219

RESULT 3  
US-08-231-193A-57/c  
; Sequence 57, Application US/08231193A  
; Patent No. 5849835  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231.193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0082  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 485..4495  
US-08-231-193A-57  
Query Match 71.7%; Score 17.2; DB 2; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGCTGGTACGTCGCATC 22

Db 2635 GGAGCGCTGGGACGGTCCCA 2614

RESULT 4  
US-08-486-273A-57/c  
; Sequence 57, Application US/08486273A  
; Patent No. 5985586  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA  
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,273A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 485..4495  
US-08-486-273A-57

Query Match 71.7%; Score 17.2; DB 2; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGCTGGTACGTCGCATC 22  
Db 2635 GGAGCGCTGGGACGGTCCCA 2614

RESULT 5  
US-08-940-086A-57/c  
; Sequence 57, Application US/08940086A  
; Patent No. 6111091  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun



;; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
;; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
;; NUMBER OF SEQUENCES: 63  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Heller Ehrman White & McAuliffe  
;; STREET: 4250 Executive Square, 7th Floor  
;; CITY: La Jolla  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/940,086A  
;; FILING DATE: 29-SEPT-97  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/231,193  
;; FILING DATE: 20-APR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/052,449  
;; FILING DATE: 20-APR-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 24735-9383C  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 450-8400  
;; TELEFAX: (619) 450-8499  
;; INFORMATION FOR SEQ ID NO: 57:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4695 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 485..4495  
;; US-08-940-086A-57  
  
Query Match 71.7%; Score 17.2; DB 3; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 GGAGCGTGGGTACGGTCGCA 22  
Db 2635 GGAGCGTGGGTACGGTCGCA 2614  
  
RESULT 6  
US-08-940-035A-57/c  
; Sequence 57, Application US/08940035A  
; Patent No. 6316811  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 29-SEPT-97  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9383C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 450-8499  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 485..4495  
; US-08-940-086A-57

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/940,035A  
;; FILING DATE: 29-SEPT-97  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/231,193  
;; FILING DATE: 20-APR-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/052,449  
;; FILING DATE: 20-APR-1993  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6362-9383E  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-238-0999  
;; TELEFAX: 619-238-0062  
;; INFORMATION FOR SEQ ID NO: 57:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4695 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: both  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 485..4495  
;; US-08-940-035A-57  
  
Query Match 71.7%; Score 17.2; DB 4; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 GGAGCGTGGGTACGGTCGCA 22  
Db 2635 GGAGCGTGGGTACGGTCGCA 2614  
  
RESULT 7  
US-08-935-105A-57/c  
; Sequence 57, Application US/08935105A  
; Patent No. 6376660  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US 08/935,105A  
; FILING DATE: 29-SEPT-97  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 20-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1993

CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4695 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 485..4495  
US-08-935-105A-57

Query Match 71.7%; Score 17.2; DB 4; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCA 22  
|||||  
Db 2635 GGAGCGTTGGCAGGTCCTCA 2614

RESULT 8  
US-08-648-797-57/c  
Sequence 57, Application US/09648797  
Patent No. 6469142  
GENERAL INFORMATION:  
APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Lu, Chin-Chun  
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 28-Aug-2000  
APPLICATION NUMBER: US/09/648,797  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/940,086A  
FILING DATE: 29-SEPT-97  
APPLICATION NUMBER: US/08/231,193  
FILING DATE: 20-APR-1994  
APPLICATION NUMBER: US/08/052,449  
FILING DATE: 20-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 24735-9383C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 450-8400  
TELEFAX: (619) 450-8499  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4695 base pairs

TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 485..4495  
SEQUENCE DESCRIPTION: SEQ ID NO: 57:  
US-09-648-797-57

Query Match 71.7%; Score 17.2; DB 4; Length 4695;  
Best Local Similarity 86.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCA 22  
|||||  
Db 2635 GGAGCGTTGGCAGGTCCTCA 2614

RESULT 9  
US-09-386-123-57/c  
Sequence 57, Application US/09386123  
Patent No. 6521413  
GENERAL INFORMATION:  
APPLICANT: Daggett, Lorrie P.  
Ellis, Steven B.  
Liaw, Chen W.  
Liaw, Chen-Chun  
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Heller Ehrman White & McAuliffe  
STREET: 4250 Executive Square, 7th Floor  
CITY: La Jolla  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/386,123  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/486,273  
FILING DATE: 06-JUNE-95  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,193  
FILING DATE: 20-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/052,449  
FILING DATE: 20-APR-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-9383F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858-450-8403  
TELEFAX: 858-587-5360  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4695 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 485..4495

US-09-386-123-57

Query Match 71.7%; Score 17.2; DB 4; Length 4695;  
Best Local Similarity 85.4%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGCGCTGGTACGTCGCA 22  
Db 2635 GGAGCGCTGGGACGCTCCA 2614

RESULT 10

US-09-198-452A-1/c

; Sequence 1, Application US/09198452A

; Patent No. 6559294

; GENERAL INFORMATION:

; APPLICANT: Griffiths, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention

; TITLE OF INVENTION: and treatment of infection

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/09/198,452A

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 1

; LENGTH: 1230025

; TYPE: DNA

; ORGANISM: Chlamydia pneumoniae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(15000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (15001)..(30000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (30001)..(45000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (45001)..(60000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (60001)..(75000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (75001)..(90000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (90001)..(105000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (105001)..(120000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (120001)..(135000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (135001)..(150000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (150001)..(165000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (165001)..(180000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (180001)..(195000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (195001)..(210000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (210001)..(225000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (225001)..(240000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (240001)..(255000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (255001)..(270000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (270001)..(285000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (285001)..(300000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (30001)..(315000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (315001)..(330000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (330001)..(345000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (345001)..(360000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (360001)..(375000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (375001)..(390000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (390001)..(405000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (405001)..(420000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (420001)..(435000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (435001)..(450000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (450001)..(465000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (465001)..(480000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (480001)..(495000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (495001)..(510000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (510001)..(525000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (525001)..(540000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (540001)..(555000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (555001)..(570000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (570001)..(585000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (585001)..(600000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (60001)..(615000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (615001)..(630000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (630001)..(645000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (645001)..(660000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (660001)..(675000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (675001)..(690000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (690001)..(705000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (705001)..(720000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (720001)..(735000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (735001)..(750000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (750001)..(765000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (765001)..(780000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (780001)..(795000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (795001)..(810000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (810001)..(825000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (825001)..(840000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (840001)..(855000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (855001)..(870000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (870001)..(885000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (885001)..(900000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (900001)..(915000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (915001)..(930000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (930001)..(945000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (945001)..(960000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (960001)..(975000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (975001)..(990000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (990001)..(1000000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1000001)..(1015000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1015001)..(1030000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1030001)..(1045000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1045001)..(1060000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1060001)..(1075000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1075001)..(1090000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1090001)..(1105000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1105001)..(1120000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1120001)..(1135000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1135001)..(1150000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1150001)..(1165000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1165001)..(1180000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1180001)..(1195000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1195001)..(1210000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1210001)..(1225000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1225001)..(1240000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1240001)..(1255000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1255001)..(1270000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1270001)..(1285000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1285001)..(1300000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1300001)..(1315000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc feature

; LOCATION: (1315001)..(1330000)

; OTHER INFORMATION: n=a or c or g or t

; NAME/KEY: misc

NAME/KEY: misc feature  
LOCATION: (585001)..(600000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (600001)..(615000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (615001)..(630000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (630001)..(645000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (645001)..(660000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (660001)..(675000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (675001)..(690000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (690001)..(705000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (705001)..(720000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (720001)..(735000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (735001)..(750000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (750001)..(765000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (765001)..(780000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (780001)..(795000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (795001)..(810000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (810001)..(825000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (825001)..(840000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (840001)..(855000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (855001)..(870000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (870001)..(885000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (885001)..(900000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (900001)..(915000)  
OTHER INFORMATION: n=a or c or g or t  
NAME/KEY: misc feature

Query Match 71.7%; Score 17.2; DB 4; Length 1230025;  
Best Local Similarity 86.4%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
1 GGAGGCGTTGGTACGGTCGCA 22

Db 447672 GGAGGCTTAGGTAGGTGCGCA 447651

## RESULT 11

US-09-335-409-1  
; Sequence 1, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-335-409-1

Query Match 69.2%; Score 16.6; DB 3; Length 68750;  
Best Local Similarity 82.6%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGAGGCGTTGGTACGGTCGCA 23

Db 4350 GGCGCGTGGGGACGGTCGCT 4372

## RESULT 12

US-09-568-102-1  
; Sequence 1, Application US/09568102  
; Patent No. 6346404  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,102  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-102-1

Query Match 69.2%; Score 16.6; DB 4; Length 68750;  
Best Local Similarity 82.6%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGAGGCGTTGGTACGGTCGCA 23

Db 4350 GGCGCGTGGGGACGGTCGCT 4372

## RESULT 13

US-09-567-969-1  
; Sequence 1, Application US/09567969

Patent No. 6355457  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,969  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-567-969-1

Query Match 69.2%; Score 16.6; DB 4; Length 68750;  
Best Local Similarity 82.6%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GGAGCGGTGGGTACGGTCGCAT 23  
Db 4350 GCGGCGGTGGGGACGGTCGCGT 4372

RESULT 14  
US-09-568-480-1  
; Sequence 1, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,480  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-480-1

Query Match 69.2%; Score 16.6; DB 4; Length 68750;  
Best Local Similarity 82.6%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GGAGCGGTGGGTACGGTCGCAT 23  
Db 4350 GCGGCGGTGGGGACGGTCGCGT 4372

RESULT 15  
US-09-568-486-1  
; Sequence 1, Application US/09568486  
; Patent No. 6355459  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-1

Query Match 69.2%; Score 16.6; DB 4; Length 68750;  
Best Local Similarity 82.6%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GGAGCGGTGGGTACGGTCGCAT 23  
Db 4350 GCGGCGGTGGGGACGGTCGCGT 4372

Search completed: April 29, 2004, 11:46:19  
Job time : 60.3529 secs





```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7570
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-10-282-122A-7570

Query Match 74.2%; Score 17.8; DB 13; Length 1734;
Best Local Similarity 90.5%; Pred. No. 81;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGGCGTTGGGTACGGTCGCA 22
Db 685 GTGGCGTTGGGTACGGTAGCA 705

RESULT 3
US-10-282-122A-37103
; Sequence 37103, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37103
; LENGTH: 1748
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-37103

Query Match 74.2%; Score 17.8; DB 13; Length 1748;
Best Local Similarity 90.5%; Pred. No. 81;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGGCGTTGGGTACGGTCGCA 22
Db 677 GTGGCGTTGGGTACGGTCGCA 697

RESULT 4
US-10-050-686A-1/c
; Sequence 1, Application US/10050686A
; Publication No. US20030077288A1
; GENERAL INFORMATION:
; APPLICANT: GOLDBERG, ALFRED L.
; APPLICANT: GOMES, MARCELO D.
; APPLICANT: LECKER, STEWART H.
; APPLICANT: JAGOE, R. THOMAS
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF MUSCLE
; FILE REFERENCE: HMV-070.01
; CURRENT APPLICATION NUMBER: US/10/050,686A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 60/262,090
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2067
; TYPE: DNA
; ORGANISM: Mus sp.
US-10-050-686A-1

Query Match 73.3%; Score 17.6; DB 15; Length 2067;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGGTACGGTCGATC 24
Db 133 GCAGGCGTTGGGTACGGTCGCTC 110

RESULT 5
```



US-10-311-455-1341  
; Sequence 1341, Application US/10311455  
; Publication No. US20030143608A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311.455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1341  
; LENGTH: 9817  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; NAME/KEY: unsure  
; LOCATION: 2636, 7012  
; OTHER INFORMATION: n is a or g or c or t  
US-10-311-455-1341  
Query Match 73.3%; Score 17.6; DB 15; Length 9817;  
Best Local Similarity 83.3%; Pred. No. 92;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 GGAGCGTTGGTACGGTCATC 24  
Db 5249 GGAGCGTTGGTACGGTCATC 5272  
RESULT 6  
US-10-282-122A-41560  
; Sequence 41560, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282.122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 41560  
; LENGTH: 1734  
; TYPE: DNA  
; ORGANISM: Versinia pestis  
US-10-282-122A-41560  
Query Match 71.7%; Score 17.2; DB 13; Length 1734;  
Best Local Similarity 86.4%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GGAGCGTTGGTACGGTCGCA 22  
Db 684 GGAGCGTTGGTACGGTCGCA 705  
RESULT 7  
US-09-841-132-381/c  
; Sequence 381, Application US/09841132  
; Patent No. US20020061848A1  
; GENERAL INFORMATION:  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Probst, Peter  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C8  
; CURRENT APPLICATION NUMBER: US/09/841,132  
; CURRENT FILING DATE: 2001-04-23  
; NUMBER OF SEQ ID NOS: 599  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 381  
; LENGTH: 1995  
; TYPE: DNA  
; ORGANISM: Chlamydia pneumoniae  
US-09-841-132-381  
Query Match 71.7%; Score 17.2; DB 9; Length 1995;  
Best Local Similarity 86.4%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GGAGCGTTGGTACGGTCGCA 22  
Db 1860 GGAGCGTTGGTACGGTCGCA 1839  
RESULT 8  
US-10-282-122A-18612/c  
; Sequence 18612, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 18612

; LENGTH: 1995

; TYPE: DNA

; ORGANISM: Chlamydia pneumoniae

US-10-282-122A-18612

Query Match

Best Local Similarity 71.7%; Score 17.2; DB 13; Length 1995;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGGTACGGTCGCA 22

|||||

DB 1860 GGAGCGTTAGGTAAGTCGCA 1839

|||||

RESULT 9

US-03-945-901-57/c

; Sequence 57, Application US/09945901

; Patent No. US20020161215A1

; GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Ellis, Steven B.

Lu, Chin-Chun

STATE: CA

COUNTRY: U.S.A.

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/945,901

FILING DATE: 24-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/940,035

FILING DATE: <Unknown>

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 18612

; LENGTH: 1995

; TYPE: DNA

; ORGANISM: Chlamydia pneumoniae

US-10-282-122A-18612

Query Match

Best Local Similarity 71.7%; Score 17.2; DB 13; Length 1995;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGGTACGGTCGCA 22

|||||

DB 1860 GGAGCGTTAGGTAAGTCGCA 1839

|||||

RESULT 9

US-03-945-901-57/c

; Sequence 57, Application US/09945901

; Patent No. US20020161215A1

; GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Ellis, Steven B.

Lu, Chin-Chun

STATE: CA

COUNTRY: U.S.A.

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/945,901

FILING DATE: 24-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/940,035

FILING DATE: <Unknown>

; APPLICATION NUMBER: US 08/052,449

; FILING DATE: 20-APR-1993

; ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-9383E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

; INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 4695 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 485..4495

SEQUENCE DESCRIPTION: SEQ ID NO: 57:

US-09-945-901-57

Query Match

Best Local Similarity 71.7%; Score 17.2; DB 9; Length 4695;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGGTACGGTCGCA 22

|||||

DB 2635 GGAGCGTTGGGACGGTCCCA 2614

|||||

RESULT 10

US-10-007-747-57/c

; Sequence 57, Application US/10007747

; Publication No. US20020161193A1

; GENERAL INFORMATION:

APPLICANT: Daggett, Lorrie P.

Ellis, Steven B.

Lu, Chin-Chun

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,747

FILING DATE: 07-Dec-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/648,797

FILING DATE: 28-Aug-2000

APPLICATION NUMBER: US/08/940,086A

FILING DATE: 29-SEPT-97

APPLICATION NUMBER: US 08/231,193

FILING DATE: 20-APR-1994

APPLICATION NUMBER: US 08/052,449

FILING DATE: 20-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Seidman, Stephanie

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 24735-9383C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 450-8400

```

; TELEFAX: (619) 450-8499
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4695 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
; MOLECULE TYPE: cdna
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 485..4495
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-007-747-57

Query Match          71.7%; Score 17.2; DB 14; Length 4695;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCA 22
DB 2635 GGAGCGTTGGGACGGTCCCA 2614

RESULT 11
US-10-038-937-57/c
; Sequence 57, Application US/10038937
; Publication No. US20030013866A1
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/038,937
; FILING DATE: 18-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIORITY INFORMATION DATA:
; APPLICATION NUMBER: 08/935,105
; FILING DATE: 29-SEPT-97
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4695 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
; MOLECULE TYPE: cdna
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 485..4495

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-038-937-57

Query Match          71.7%; Score 17.2; DB 15; Length 4695;
Best Local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCGCA 22
DB 2635 GGAGCGTTGGGACGGTCCCA 2614

RESULT 12
US-09-997-722-292
; Sequence 292, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 292
; LENGTH: 96588
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-997-722-292

Query Match          71.7%; Score 17.2; DB 12; Length 96588;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGCGTTGGTACGGTCGCAT 23
DB 61567 GAGCGTTGGTACGGTCGCAT 61588

RESULT 13
US-10-289-762-1/c
; Sequence 1, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t

```



LOCATION: (585001)..(600000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (600001)..(615000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (615001)..(630000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (630001)..(645000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (645001)..(660000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (660001)..(675000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (675001)..(690000)  
OTHER INFORMATION: n=a or c or g or t  
FEATURE:

Query Match 71.7%; Score 17.2; DB 16; Length 1230025;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGGTACGGTCGCA 22  
|||||  
Db 447672 GGAGCGTTAGTAAGTCGCA 447651

## RESULT 14

US-10-369-493-39062/c  
Sequence 39062, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39062  
LENGTH: 933  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39062

Query Match 69.2%; Score 16.6; DB 16; Length 933;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GAGCGTTGGGTACGGTCGATC 24  
|||||  
Db 434 GAGCGTTGGAGACGATCGCTTC 412

## RESULT 15

US-10-369-493-39429/c  
Sequence 39429, Application US/10369493

Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 39429  
LENGTH: 933  
TYPE: DNA  
ORGANISM: Xanthomonas campestris  
US-10-369-493-39429

Query Match 69.2%; Score 16.6; DB 16; Length 933;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GAGCGTTGGGTACGGTCGATC 24  
|||||  
Db 434 GAGCGTTGGAGACGATCGCTTC 412

Search completed: April 29, 2004, 20:44:41  
Job time : 459 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 198.118 Seconds  
(without alignments)  
514.627 Million cell updates/sec

Title: US-10-624-714-7

Perfect score: 24  
Sequence: 1 gggagcgttggtacggtgcac 24

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04.\*  
1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002s.\*  
7: geneseqn2003as.\*  
8: geneseqn2003bs.\*  
9: geneseqn2003cs.\*  
10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	1097	2	AAT86594
2	24	100.0	1097	2	AAT76948
3	24	100.0	86114	6	ABX09143
4	24	100.0	110000	4	AA199682_05
5	24	100.0	110000	4	AA199683_05
6	18.2	75.8	1207	6	ABQ29098
7	18.2	75.8	1207	6	ABQ29099
8	18.2	75.8	1207	9	ABQ29098
9	18.2	75.8	1207	9	ABQ29099
10	17.8	74.2	787	6	ABQ50997
11	17.8	74.2	787	6	ABQ50997
12	17.8	74.2	1734	7	ACA19700
13	17.8	74.2	1734	7	ACA19700
14	17.6	73.3	2067	9	ACA49233
15	17.6	73.3	9817	6	ABQ29098
16	17.6	73.3	110000	4	AA199682_12
17	17.6	73.3	110000	4	AA199683_12
18	17.2	71.7	1734	7	ACA53690
19	17.2	71.7	1995	6	ABL92608
20	17.2	71.7	1995	6	ACA30742
21	17.2	71.7	4695	2	AAQ79378
22	17.2	71.7	4695	2	AAV82917
23	17.2	71.7	4695	3	AAZ38731

c	24	17.2	71.7	4695	3	AAA95040	Aaa95040 Human N-m
c	25	17.2	71.7	4695	6	AB199196	Ab199196 Human NMD
c	26	17.2	71.7	4695	6	AA147384	Aa147384 Human NMD
c	27	17.2	71.7	4695	7	ABX98543	Abx98543 Human N-m
c	28	17.2	71.7	4695	7	ABX92892	Abx92892 Human N-m
c	29	17.2	71.7	4695	7	ABX77688	Abx77688 DNA encod
c	30	17.2	71.7	4695	8	ACD98353	AcD98353 Human NMD
c	31	17.2	71.7	4695	9	AAU59685	AaU59685 Human NMD
c	32	17.2	71.7	96588	8	ADA03026	Ada03026 Human MBN
c	33	17.2	71.7	96588	9	ADB72764	AdB72764 Human MBN
c	34	17.2	71.7	96588	9	ADC85506	AdC85506 Human MBN
c	35	17.2	71.7	110000	2	AAK91990_04	Continuation (5 of
c	36	17.2	71.7	273254	3	AAK81914	AaK81914 Chlamydia
c	37	17	70.8	842	6	ABQ13524	AbQ13524 Oligonuc
c	38	17	70.8	842	6	ABQ13525	AbQ13525 Oligonuc
c	39	16.8	70.0	539	6	ABQ47534	AbQ47534 Oligonuc
c	40	16.8	70.0	539	6	ABQ47535	AbQ47535 Oligonuc
c	41	16.8	70.0	996	6	ABQ42151	AbQ42151 Oligonuc
c	42	16.8	70.0	996	6	ABQ42150	AbQ42150 Oligonuc
c	43	16.8	70.0	2501	9	AD854102	Ad854102 Pretreat
c	44	16.8	70.0	2501	9	AD884094	Ad884094 Human lym
c	45	16.8	70.0	3628	7	ABZ09972	AbZ09972 Haematopo

#### ALIGNMENTS

RESULT 1  
AAT86594  
ID AAT86594 standard; DNA; 1097 BP.

AC AAT86594;

DT 14-OCT-1998 (first entry)

XX Mycobacterial heparin-binding haemagglutinin antigen gene.

KW Surface protein; Mycobacterium bovis; BCG; adhesion; epithelium; ds;  
KW Bacille Calmette-Guerin; Mycobacterium tuberculosis; epithelial cell;  
KW heparin-binding haemagglutinin antigen; PCR; primer; amplification;  
KW probe; hybridisation; chromosome; vaccine; diagnosis; immunoassay.

XX Mycobacterium bovis.

OS Mycobacterium bovis.

FH Key Location/Qualifiers

FT CDS 331..930

FT /\*tag= a

FT /product= "HBHA"

FT /note= "heparin-binding haemagglutinin antigen"

XX FR2748749-A1.

XX 21-NOV-1997.

XX 17-MAY-1996; 96FR-00006169.

XX 17-MAY-1996; 96FR-00006169.

XX (INSP ) INST PASTEUR LILLE.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Menozzi F, Loch C;

XX WPI; 1998-021392/03.

XX P-PSDB; AAW43082.

XX Mycobacterial heparin-binding haemagglutinin polypeptide - useful for

XX vaccination against and diagnosis of mycobacterial infections.

XX Claim 19; Fig 10; 50pp; French.

XX This sequence represents the gene encoding a surface protein found on

XX Mycobacterium bovis BCG (Bacille Calmette-Guerin) or M. tuberculosis,

CC which enables mycobacteria to adhere to host, especially epithelial,  
 CC cells. The protein was isolated by passing a culture of BCG  
 CC microorganisms over a Heparin-Sepharose column and eluting proteins with  
 CC a gradient of 0-0.5 M NaCl. Fractions were separated by gel  
 CC electrophoresis and a 28 kD heparin binding protein purified. N-terminal  
 CC and internal peptide fractions were sequenced and the amino acid  
 CC sequences used to design PCR primers (AAT86596-R86599). These amplified a  
 CC 150 bp fragment (AAT86595) used as a probe to isolate chromosomal  
 CC fragments containing the gene which encodes a heparin-binding  
 CC haemagglutinin antigen (HBHA). The protein or peptides derived from it  
 CC (e.g. AAW43081), can be used in vaccines against mycobacterial infections  
 CC (particularly *M. bovis* or *M. tuberculosis*), or can be used to diagnose  
 CC mycobacterial infection (by immunoassay detection of anti-HBHA  
 CC antibodies)  
 XX  
 SQ Sequence 1097 BP; 236 A; 324 C; 365 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 2; Length 1097;  
 Best Local Similarity 100.0%; Pred. No. 0.23; 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGAGGCGTTGGGTACGGTCGCATC 24  
 Db 732 GGAGGCGTTGGGTACGGTCGCATC 755

RESULT 2  
 AAT76948  
 ID AAT76948 standard; DNA; 1097 BP.  
 AC AAT76948;  
 DT 14-OCT-1998 (first entry)  
 DE Mycobacterial heparin-binding haemagglutinin antigen gene.  
 KW Surface protein; Mycobacterium bovis; BCG; adhesion; epithelium; ds;  
 KW Bacille Calmette-Guerin; Mycobacterium tuberculosis; epithelial cell;  
 KW heparin-binding haemagglutinin antigen; PCR; primer; amplification;  
 KW probe; hybridisation; chromosome; vaccine; diagnosis; immunoassay.  
 XX  
 OS Mycobacterium bovis.

Key Location/Qualifiers  
 CDS 331..930  
 /tag= a  
 /product= "HBHA"  
 /note= "heparin-binding haemagglutinin antigen"  
 repeat\_region 811..864  
 /tag= b  
 /rpt\_unit= 811..828  
 repeat\_region 865..915  
 /tag= d  
 /rpt\_unit= 865..885

WO9744463-A2.  
 XX  
 XX 27-NOV-1997.  
 XX  
 XX 20-MAY-1997; 97WO-FR000886.  
 XX  
 XX 17-MAY-1996; 96FR-00006168.  
 XX (INSP ) INST PASTEUR LILLE.  
 XX PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 XX  
 XX Menozzi F, Loch C;  
 XX  
 XX WPI; 1998-018517/02.  
 XX P-PSDB; AAW44936.  
 XX  
 XX New peptide(s) involved in adhesion of mycobacteria to epithelial cells -  
 XX used in vaccines and for diagnosing mycobacterial infection, also use of

PT sulphated saccharide(s) and glyco-conjugate(s) to prevent mycobacterial  
 PT adhesion.  
 XX  
 XX Claim 21; Fig 10; 52pp; French.  
 XX  
 CC This sequence represents the gene encoding a surface protein found on  
 CC Mycobacterium bovis BCG (Bacille Calmette-Guerin) or *M. tuberculosis*,  
 CC which enables mycobacteria to adhere to host, especially epithelial,  
 CC cells. The protein was isolated by passing a culture of BCG  
 CC microorganisms over a Heparin-Sepharose column and eluting proteins with  
 CC a gradient of 0-0.5 M NaCl. Fractions were separated by gel  
 CC electrophoresis and a 28 kD heparin binding protein purified. N-terminal  
 CC and internal peptide fractions were sequenced and the amino acid  
 CC sequences used to design PCR primers (AAT76949-R76952). These amplified a  
 CC 150 bp fragment (AAT76947) used as a probe to isolate chromosomal  
 CC fragments containing the gene which encodes a heparin-binding  
 CC haemagglutinin antigen (HBHA). The protein or peptides derived from it  
 CC (e.g. AAW44934), can be used in vaccines against mycobacterial infections  
 CC (particularly *M. bovis* or *M. tuberculosis*), or can be used to diagnose  
 CC mycobacterial infection (by immunoassay detection of anti-HBHA  
 CC antibodies)  
 XX  
 SQ Sequence 1097 BP; 236 A; 324 C; 365 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 2; Length 1097;  
 Best Local Similarity 100.0%; Pred. No. 0.23; 0; Indels 0; Gaps 0;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGAGGCGTTGGGTACGGTCGCATC 24  
 Db 732 GGAGGCGTTGGGTACGGTCGCATC 755

RESULT 3  
 ABX09143  
 ID ABX09143 standard; DNA; 86114 BP.  
 AC ABX09143;  
 XX  
 XX 08-APR-2003 (first entry)  
 XX  
 DE Mycobacterium tuberculosis H37Rv BAC clone BAC-RV265.  
 XX  
 KW Mycobacteriosis; survival; virulence; protective antigen; vaccine;  
 KW mycobacterial disease; tuberculosis; leprosy; ds; cosmid.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 XX WO200274903-A2.  
 XX  
 XX 26-SEP-2002.  
 XX  
 XX 22-FEB-2002; 2002WO-IB001973.  
 XX  
 XX 22-FEB-2001; 2001US-0270123P.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX  
 XX Cole S;  
 XX  
 XX WPI; 2002-759885/82.

Identifying and selecting genes for survival or virulence of mycobacteria  
 by a comparative genomic analysis of the sequences of Mycobacterium  
 tuberculosis and *M. leprae*.  
 XX  
 XX Disclosure; Fig 7; 874pp; English.  
 XX  
 CC This invention relates to a novel method for identifying essential genes  
 CC for survival or virulence of mycobacteria species. The method comprises  
 CC aligning the genomic sequence of a first mycobacterium species on a  
 CC genomic sequence of a second mycobacterium species and selecting a  
 CC polynucleotide sequence that is highly conserved in both genomes with no



CC counterparts in other bacterial genomic sequences and that corresponds to  
CC an essential gene for the survival or virulence of mycobacterium species.  
CC The method of the invention is useful for detecting M. tuberculosis or M.  
CC leprae infection. The method reduces the number of potential new targets  
CC and protective antigens for new drugs and vaccine compositions to treat  
CC and prevent mycobacterial diseases, particularly tuberculosis and  
CC leprosy. The present sequence represents a Mycobacterium cosmid DNA  
CC sequence used in the method of the invention

SQ Sequence 86114 BP; 14693 A; 27103 C; 28984 G; 15334 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 6; Length 86114;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGTACGTCGCATC 24

Db 52795 GGAGCGTTGGTACGTCGCATC 52818

RESULT 4  
AAI99682\_05  
Continuation (6 of 45) of AAI99682 from base 500001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_01	1	110000
WP	AAI99682_02	100001	210000
WP	AAI99682_03	200001	310000
WP	AAI99682_04	300001	410000
WP	AAI99682_05	400001	510000
WP	AAI99682_06	500001	610000
WP	AAI99682_07	600001	710000
WP	AAI99682_08	700001	810000
WP	AAI99682_09	800001	910000
WP	AAI99682_10	900001	1010000
WP	AAI99682_11	1000001	1110000
WP	AAI99682_12	1100001	1210000
WP	AAI99682_13	1200001	1310000
WP	AAI99682_14	1300001	1410000
WP	AAI99682_15	1400001	1510000
WP	AAI99682_16	1500001	1610000
WP	AAI99682_17	1600001	1710000
WP	AAI99682_18	1700001	1810000
WP	AAI99682_19	1800001	1910000
WP	AAI99682_20	1900001	2010000
WP	AAI99682_21	2000001	2110000
WP	AAI99682_22	2100001	2210000
WP	AAI99682_23	2200001	2310000
WP	AAI99682_24	2300001	2410000
WP	AAI99682_25	2400001	2510000
WP	AAI99682_26	2500001	2610000
WP	AAI99682_27	2600001	2710000
WP	AAI99682_28	2700001	2810000
WP	AAI99682_29	2800001	2910000
WP	AAI99682_30	2900001	3010000
WP	AAI99682_31	3000001	3110000
WP	AAI99682_32	3100001	3210000
WP	AAI99682_33	3200001	3310000
WP	AAI99682_34	3300001	3410000
WP	AAI99682_35	3400001	3510000
WP	AAI99682_36	3500001	3610000
WP	AAI99682_37	3600001	3710000
WP	AAI99682_38	3700001	3810000
WP	AAI99682_39	3800001	3910000
WP	AAI99682_40	3900001	4010000
WP	AAI99682_41	4000001	4110000
WP	AAI99682_42	4100001	4210000
WP	AAI99682_43	4200001	4310000
WP	AAI99682_44	4300001	4411529

Query Match 100.0%; Score 24; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.25;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGTACGTCGCATC 24  
Db 66196 GGAGCGTTGGTACGTCGCATC 66219

RESULT 5  
AAI99683\_05  
Continuation (6 of 44) of AAI99683 from base 500001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP	Fragment Name	Begin	End
WP	AAI99683_01	1	110000
WP	AAI99683_02	100001	210000
WP	AAI99683_03	200001	310000
WP	AAI99683_04	300001	410000
WP	AAI99683_05	400001	510000
WP	AAI99683_06	500001	610000
WP	AAI99683_07	600001	710000
WP	AAI99683_08	700001	810000
WP	AAI99683_09	800001	910000
WP	AAI99683_10	900001	1010000
WP	AAI99683_11	1000001	1110000
WP	AAI99683_12	1100001	1210000
WP	AAI99683_13	1200001	1310000
WP	AAI99683_14	1300001	1410000
WP	AAI99683_15	1400001	1510000
WP	AAI99683_16	1500001	1610000
WP	AAI99683_17	1600001	1710000
WP	AAI99683_18	1700001	1810000
WP	AAI99683_19	1800001	1910000
WP	AAI99683_20	1900001	2010000
WP	AAI99683_21	2000001	2110000
WP	AAI99683_22	2100001	2210000
WP	AAI99683_23	2200001	2310000
WP	AAI99683_24	2300001	2410000
WP	AAI99683_25	2400001	2510000
WP	AAI99683_26	2500001	2610000
WP	AAI99683_27	2600001	2710000
WP	AAI99683_28	2700001	2810000
WP	AAI99683_29	2800001	2910000
WP	AAI99683_30	2900001	3010000
WP	AAI99683_31	3000001	3110000
WP	AAI99683_32	3100001	3210000
WP	AAI99683_33	3200001	3310000
WP	AAI99683_34	3300001	3410000
WP	AAI99683_35	3400001	3510000
WP	AAI99683_36	3500001	3610000
WP	AAI99683_37	3600001	3710000
WP	AAI99683_38	3700001	3810000
WP	AAI99683_39	3800001	3910000
WP	AAI99683_40	3900001	4010000
WP	AAI99683_41	4000001	4110000
WP	AAI99683_42	4100001	4210000
WP	AAI99683_43	4200001	4310000
WP	AAI99683_44	4300001	4403765

Query Match 100.0%; Score 24; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAGCGTTGGTACGTCGCATC 24  
Db 67638 GGAGCGTTGGTACGTCGCATC 67661

RESULT 6  
ABQ29098  
ID ABQ29098 standard; DNA; 1207 BP.

XX AC  
XX AC  
DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 15689.  
DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP010074.  
XX 01-SEP-2000; 2000DE-01043826.  
XX 05-SEP-2000; 2000DE-01044543.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used;  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX Sequence 1207 BP; 184 A; 124 C; 420 G; 479 T; 0 U; 0 Other;  
SQ Query Match 75.8%; Score 18.2; DB 6; Length 1207;  
Best Local Similarity 87.0%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGAGCGCTGGTACGTCGCAT 23  
Db 179 GTAGCGCTGGTACGTCGCAT 201  
RESULT 7  
ID ABQ29099/c  
XX ABQ29099 standard; DNA; 1207 BP.  
XX ABQ29099;  
XX 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 15690.  
XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP010074.  
XX 01-SEP-2000; 2000DE-01043826.  
XX 05-SEP-2000; 2000DE-01044543.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used;  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX Sequence 1207 BP; 479 A; 420 C; 124 G; 184 T; 0 U; 0 Other;  
SQ Query Match 75.8%; Score 18.2; DB 6; Length 1207;  
Best Local Similarity 87.0%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGAGCGCTGGTACGTCGCAT 23  
Db 1029 GTAGCGCTGGTACGTCGCAT 1007  
RESULT 8  
ID ADE72578/c  
XX ADE72578 standard; DNA; 2120 BP.  
XX ADE72578;  
XX 29-JAN-2004 (first entry)  
XX Human endometrial specific gene, SEQ ID NO 18.  
XX cytostatic; vaccine; human; endometrial specific genes;  
KW endometrial specific protein; endometrial cancer; ds.  
XX



PT amphotericin.  
 XX  
 PS Claim 1; Page 52-114; 276pp; English.  
 XX The invention relates to the gene cluster encoding the polypeptides responsible for the biosynthesis of the polyene antibiotic amphotericin (amph) of *Streptomyces nodosus*. Polynucleotides of the invention are useful for preparing amphotericin derivatives or analogue antibiotic agents with altered properties and in the biosynthesis of polypeptides other than amphotericin. amphIII, amphII or amphDI mutants are useful for producing amphotericin derivatives glycosylated with alternative sugars; amphIII or amphDI gene sequences are useful in engineered biosynthesis of perosaminyl-amphoteronolide B; amphDII or amphDII and amphN gene sequences are useful in the engineered biosynthesis of perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDII, amphDII and amphDI gene sequences are useful for preparing polypeptides capable of addition of mycosamine to a polyketide other than amphoteronolide A or B or for preparing polypeptides for in vitro synthesis of GPP-mycosamine. The present sequence is *S. nodosus* amph biosynthetic gene cluster

XX  
 SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;  
 Query Match 75.8%; Score 18.2; DB 7; Length 113193;  
 Best Local Similarity 87.0%; Pred. No. 1.1e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCCGAT 23  
 |||||  
 DB 66497 GGAGCGTTGGTACGGTCCGCT 66475

RESULT 10  
 ABQ50996  
 ID ABQ50996 standard; DNA; 787 BP.  
 AC ABQ50996;  
 XX  
 DT 12-JUL-2002 (first entry)  
 DE  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 37587.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP010074.  
 XX  
 PF 01-SEP-2000; 2000DE-01043826.  
 PR  
 PR 05-SEP-2000; 2000DE-01044543.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 XX WPI; 2002-371829/40.  
 XX  
 XX Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.  
 XX  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNPs); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the CC disclosure of the invention

XX  
 SQ Sequence 787 BP; 86 A; 121 C; 316 G; 264 T; 0 U; 0 Other;  
 Query Match 74.2%; Score 17.8; DB 6; Length 787;  
 Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGCGTTGGTACGGTCCG 21  
 |||||  
 DB 615 GGAGCGTTAGGTACGGTCCG 635

RESULT 11  
 ABQ50997/c  
 ID ABQ50997 standard; DNA; 787 BP.  
 XX  
 AC ABQ50997;  
 XX  
 DT 12-JUL-2002 (first entry)  
 DE  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 37588.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP010074.  
 XX  
 PF 01-SEP-2000; 2000DE-01043826.  
 PR  
 PR 05-SEP-2000; 2000DE-01044543.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 XX WPI; 2002-371829/40.  
 XX  
 XX Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.  
 XX  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the



FT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 37103; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 621.3 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: the sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1748 BP; 362 A; 431 C; 556 G; 399 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 7; Length 1748;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGCGTGTGGTACGGTCGCA 22

Db 677 GTGGCGTGTGGACGGTCGCA 697

RESULT 14

AD24728/c  
ID AD24728 standard; cDNA; 2067 BP.

XX AC AD24728;

XX DT 29-JAN-2004 (first entry)

XX DE Mouse cDNA encoding F-box protein, Atrophin-1.

XX KW Mouse; F-box protein; Atrophin-1; ss; gene; ubiquitination; SCF complex;  
KW ubiquitin-protein ligase; muscle wasting disorder;  
KW muscle stem cell proliferation; cachexia; rheumatoid arthritis;  
KW rheumatoid spondylitis; osteoarthritis; gouty arthritis; sepsis;  
KW septic shock; endotoxic shock; gram-negative sepsis;  
KW toxic shock syndrome; adult respiratory distress syndrome;  
KW cerebral malaria; chronic pulmonary inflammatory disease;  
KW prolonged inactivity; microgravity environment; muscle mass maintenance;  
KW protein degradation; cell proliferation; cell differentiation;  
KW cell survival.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
XX CDS 328..1395  
/\*tag= a

/product= "Atrophin-1"

US2003077288-A1.

XX 24-APR-2003.

XX 16-JAN-2002; 2002US-00050696.

XX 16-JAN-2001; 2001US-0262090P.

PA (GOLD/) GOLDBERG A L.

PA (GOME/) GOMES M D.

PA (LECK/) LECKER S H.

PA (JAGO/) JAGOE R T.

PI Goldberg AL, Gomes MD, Lecker SH, Jagoe RT;

XX WPI: 2003-556144/52.

XX P-PSDB; AD24729.

XX Novel isolated and/or recombinant cell- or tissue-specific F-box protein  
XX useful for modulating protein degradation, cell proliferation, cell  
XX differentiation and/or cell survival by modulating protein  
XX ubiquitination.

XX Claim 15; Fig 5A; 42pp; English.

XX The invention relates to an isolated and/or recombinant cell- or tissue-  
XX specific F-box protein, especially muscle expressed Atrophin-1. Also  
XX included are an isolated nucleic acid or its complementary sequence  
XX comprising a nucleotide sequence encoding Atrophin-1, an expression  
XX vector capable of replicating in a prokaryotic or eukaryotic cell  
XX (comprising the Atrophin-1 nucleic acid) and a transcriptional regulatory  
XX sequence operably linked to the nucleotide sequence, a host cell  
XX transfected with the vector (and expressing the recombinant polypeptide),  
XX preparation of Atrophin-1, a transgenic animal having cells which harbour  
XX a transgene comprising the vector (or in which a gene comprising the  
XX nucleic acid is disrupted), an isolated nucleic acid which selectively  
XX hybridises under high stringency conditions to at least ten nucleotides  
XX of mouse atrophin-1 gene sequence (AB24728) or its complementary  
XX sequences (which nucleic acid can specifically detect or amplify a  
XX nucleic acid sequence of a vertebrate cell or tissue-specific F-box  
XX gene), a reconstituted protein mixture comprising Atrophin-1 and a  
XX substrate protein, an assay for identifying an inhibitor of cell or  
XX tissue-specific Atrophin-1-mediated ubiquitination, identifying an  
XX inhibitor of interaction between a substrate polypeptide and an SCF  
XX (ubiquitin-protein ligase) complex including Atrophin-1, creating a  
XX patient suffering from a muscle wasting disorder, maintaining or  
XX increasing muscle mass of an animal (or inhibiting protein degradation in  
XX muscle tissue of a patient without substantially affecting protein  
XX degradation in other tissues) involves administering an atrophin-1  
XX inhibitor to inhibit the expression and/or activity of atrophin-1 and  
XX stimulating the proliferation of muscle stem cells (involves contacting  
XX the stem cells with a compound capable of inhibiting the expression  
XX and/or activity of atrophin-1). Atrophin-1 is useful for diagnosing a  
XX muscle wasting disorder. The patient is suffering from cachexia, cachexia  
XX secondary to infection or malignancy, cachexia secondary to AIDS,  
XX rheumatoid arthritis, rheumatoid spondylitis, osteoarthritis, gouty  
XX arthritis, sepsis, septic shock, endotoxic shock, gram-negative sepsis,  
XX toxic shock syndrome, adult respiratory distress syndrome, cerebral  
XX malaria, chronic pulmonary inflammatory diseases, prolonged inactivity or  
XX prolonged exposure to a microgravity environment. The method further  
XX involves applying a treatment to the patient which inhibits the  
XX expression and/or activity of the F-box polypeptide. The method detailed  
XX above are useful for increasing or maintaining the muscle mass of a human  
XX or livestock animal such as cow, pig, goat or sheep. Atrophin-1 and  
XX nucleic acid are useful for modulating the protein degradation, cell  
XX proliferation, cell differentiation and/or cell survival by modulating  
XX protein ubiquitination, and thus for treating disorders associated with  
XX these functions and for diagnosing or prognosing whether a subject is at  
XX risk for developing a disorder associated with protein degradation, cell  
XX proliferation, cell differentiation and/or cell survival. The present  
XX sequence encodes mouse Atrophin-1.

Search completed: April 29, 2004, 04:57:06  
Job time : 201.218 secs

```

XX SQ Sequence 2067 BP; 572 A; 512 C; 527 G; 456 T; 0 U; 0 Other;
Query Match 73.3%; Score 17.6; DB 9; Length 2067;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GGAGCGGTGGTACGTCGCATC 24
Db 133 GCAGCGGTGGGAAGTCGCCTC 110

RESULT 15
ABL33368
ID ABL33368 standard; DNA; 9817 BP.
XX
AC ABL33368;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1341.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianemic; cytostatic; nontropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
FD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
FA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1341; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 9817 BP; 2814 A; 174 C; 2033 G; 4794 T; 0 U; 2 Other;

Query Match 73.3%; Score 17.6; DB 6; Length 9817;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GGAGCGGTGGTACGTCGCATC 24
Db 5249 GCAGCGGTGGTTTGGTCGCATC 5272

```

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 29, 2004, 02:25:04 ; Search time 392.345 Seconds  
(without alignments)  
2872.264 Million cell updates/sec

Title: US-10-624-714-8  
Perfect score: 26  
Sequence: 1 gattcgagccgactactcttgggt 26

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_ma.\*  
20: em\_ma.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hg\_hum.\*  
31: em\_hg\_inv.\*  
32: em\_hg\_other.\*  
33: em\_hg\_mus.\*  
34: em\_hg\_pin.\*  
35: em\_hg\_rod.\*  
36: em\_hg\_mam.\*  
37: em\_hg\_vrt.\*  
38: em\_sv.\*  
39: em\_hgo\_hum.\*  
40: em\_hgo\_mus.\*  
41: em\_hgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	26	100.0	1097	1	AF074390	AF074390 Mycobacte
C 2	26	100.0	1097	6	A67974	A67974 Sequence 19
C 3	26	100.0	15670	1	AE006951	AE006951 Mycobacte
C 4	26	100.0	86114	6	AX704277	AX704277 Sequence
C 5	26	100.0	324050	1	BX248335	BX248335 Mycobacte
C 6	26	100.0	342416	1	BX842573	BX842573 Mycobacte
C 7	19.6	75.4	168776	5	BX465204	BX465204 Zebrafish
C 8	19	73.1	73112	9	HS072333	HS072333 Human DNA
C 9	19	73.1	150881	2	AC072036	AC072036 Homo sapi
C 10	18.8	72.3	2187	9	BC028151	BC028151 Homo sapi
C 11	18.8	72.3	2622	9	HMLSPRO	M99578 Human lymph
C 12	18.8	72.3	3233	9	HUMXE7A	L03426 Human XE7 m
C 13	18.8	72.3	184927	2	AL683875	AL683875 Homo sapi
C 14	18.8	72.3	189825	9	AL683807	AL683807 Human DNA
C 15	18.6	71.5	1542	5	AY398405	AY398405 Danio rer
C 16	18.6	71.5	176343	9	CNS01DX3	AL139021 Human chr
C 17	18.6	71.5	176644	9	AP002982	AP002982 Homo sapi
C 18	18.6	71.5	183445	2	AC132289	AC132289 Mus muscu
C 19	18.6	71.5	201190	2	AC022256	AC022256 Homo sapi
C 20	18.6	71.5	217516	2	EX629346	EX629346 Homo sapi
C 21	18.6	71.5	220335	2	EX276106	EX276106 Danio rer
C 22	18.6	71.5	244533	2	AC123483	AC123483 Rattus no
C 23	18.6	71.5	246384	2	AC111820	AC111820 Rattus no
C 24	18.2	70.0	346	8	AB004771	AB004771 Oryza lon
C 25	18.2	70.0	1300	8	MISBAP6G	X57100 S.bicolor m
C 26	18.2	70.0	89677	9	MISBAP6M	X57101 S.bicolor m
C 27	18.2	70.0	93559	2	AC024685	AL733596 Human DNA
C 28	18.2	70.0	136932	2	AC137911	AC024685 Homo sapi
C 29	18.2	70.0	154452	2	AC137912	AC137911 Felis cat
C 30	18.2	70.0	161047	9	AL772182	AC137912 Felis cat
C 31	18.2	70.0	168147	9	AL139092	AL772182 Human DNA
C 32	18.2	70.0	175695	2	AC021717	AL139092 Human DNA
C 33	18.2	70.0	222938	2	AC095075	AC021717 Homo sapi
C 34	18.2	70.0	226225	2	AC121002	AC095075 Rattus no
C 35	18.2	70.0	300425	1	AP005044	AC121002 Rattus no
C 36	18.2	70.0	2308	8	AY393879	AP005044 Streptomy
C 37	18	69.2	13912	1	AE000975	AY393879 Metschnik
C 38	18	69.2	21581	6	AX647495	AE000975 Archaeogl
C 39	18	69.2	77378	2	AC016183	AX647495 Sequence
C 40	18	69.2	79013	9	BS000030	AC016183 Homo sapi
C 41	18	69.2	81207	9	HS12513	BS000030 Pan trogl
C 42	18	69.2	120323	2	BX539322	AL033528 Human DNA
C 43	18	69.2	138295	5	EX293559	BX539322 Danio rer
C 44	18	69.2	139145	2	AC146326	BX293559 Zebrafish
C 45	18	69.2	139145	2	AC146326	AC146326 Felis cat

ALIGNMENTS

RESULT 1  
AF074390/c  
LOCUS  
DEFINITION  
Mycobacterium tuberculosis heparin-binding hemagglutinin (hbha)  
1097 bp DNA linear BCT 13-JUL-1998  
Gene, complete cds.  
ACCESSION  
AF074390  
VERSION  
AF074390.1  
KEYWORDS  
GI:3309265  
SOURCE  
Mycobacterium tuberculosis  
ORGANISM  
Mycobacterium tuberculosis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 1097)  
REFERENCE  
Menozzi,F.D., Rouse,J.H., Alavi,M., Laude-Sharp,M., Muller,J.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES

Bischoff, R., Brennan, M.J. and Loch, C.  
Identification of a heparin-binding hemagglutinin present in  
Mycobacterium  
J. Exp. Med. 184 (3), 993-1001 (1996)  
97188915  
9064359  
2 (bases 1 to 1097)  
Menozzi, F.D., Bischoff, R., Brennan, M.J., Fort, E. and Loch, C.  
Direct Submission  
Submitted (25-JUN-1998) Microbiology, Institut Pasteur, Calmette 1,  
Lille 59019, France  
Location/Qualifiers  
1. .1097

source  
/organism="Mycobacterium tuberculosis"  
/mol\_type="genomic DNA"  
/isolate="Paris"  
/db\_xref="taxon:1773"  
331. .930  
/gene="hbha"  
331. .930  
/gene="hbha"  
/note="adhesin; HBHA"  
/codon\_start=1  
/transl\_table=11  
/product="heparin-binding hemagglutinin"  
/protein\_id="AAC26052.1"  
/db\_xref="GI:3309266"  
/translation="MAENSNIDDIKAPLLAALGAADLALATVNLINLREABETRT  
DTRSVESRRLTKLQEDLEPQLTELEKFTAEELKAAAGYLEAATSYNELVERG  
EALERLRSQSQSFVSPAGYVDQAVELTQEALGTVAQTRAVGERAAKLVGIELP  
KKAAPAKKAAAPAKKAAAPAKKAAAKKAAKVKTKQ"

Query Match 100.0%; Score 26; DB 1; Length 1097;  
Best Local Similarity 100.0%; Pred. No. 0.054;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTCTCGGT 26  
|||||  
Db 944 GATTCGGAGCCGACTACTCTCGGT 919  
|||||

## ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 1097;  
Best Local Similarity 100.0%; Pred. No. 0.054;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTCTCGGT 26  
|||||  
Db 944 GATTCGGAGCCGACTACTCTCGGT 919  
|||||

RESULT 2  
A67974/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES

Sequence 19 from Patent WO9744463.  
A67974  
A67974.1 GI:4756790  
unidentified  
unidentified  
unclassified.  
1 (bases 1 to 1097)  
Menozzi, F. and Loch, C.  
IDENTIFICATION AND CLONING OF A MYCOBACTERIAL ANTIGEN CORRESPONDING  
TO A HEPARIN-BINDING HAEMAGGLUTININ  
Patent: WO 9744463-A 19 27-NOV-1997;  
PASTEUR INSTITUT (FR)  
Other publication FR 2748748 1997/1121.  
Location/Qualifiers  
1. .1097

## CDS

source  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
331. .925  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB42304.1"  
/db\_xref="GI:4756791"  
/db\_xref="REMTREMBL:CAB42304"  
/translation="MAENSNIDDIKAPLLAALGAADLALATVNLINLREABETRT  
DTRSVESRRLTKLQEDLEPQLTELEKFTAEELKAAAGYLEAATSYNELVERG  
EALERLRSQSQSFVSPAGYVDQAVELTQEALGTVAQTRAVGERAAKLVGIELP

## ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 1097;  
Best Local Similarity 100.0%; Pred. No. 0.054;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTCTCGGT 26  
|||||  
Db 944 GATTCGGAGCCGACTACTCTCGGT 919  
|||||

RESULT 3  
AE006951/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

15670 bp DNA linear BCT 27-APR-2001  
Mycobacterium tuberculosis CDC1551, section 37 of 280 of the  
complete genome.  
AE006951 AE000516  
AE006951.1 GI:13879993  
Mycobacterium tuberculosis CDC1551  
Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 15670)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,  
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,  
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,  
Salzberg, S.L., Delcher, A., Ufferback, T., Weidman, J., Khouri, H.,  
Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains  
Unpublished  
2 (bases 1 to 15670)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,  
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,  
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,  
Salzberg, S.L., Delcher, A., Ufferback, T., Weidman, J., Khouri, H.,  
Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. .15670  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
complement(102. .542)  
/gene="MT0487"  
complement(102. .542)  
/gene="MT0487"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44710.1"  
/db\_xref="GI:13879994"  
/translation="MGAGGVEVVLASPLVGLLCTTLMGKHDKIGYDEPLGIRTLPLV  
LIGETCATVTLAMVGVLLIIVNVMLAMPFCWSPGRCPCGRKCGPISCDGGPS  
SRHPRGRCRCNRPWPGCTCVRVRCVLMWRSVFGAPGDFR"  
complement(473. .961)  
/gene="MT0488"  
complement(473. .961)  
/gene="MT0488"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44711.1"  
/db\_xref="GI:13879995"

gene  
CDS  
/translation="MPDAGAGSELRSWAYALRTNPPADGPTDTVTRWLVVVTRAAVLP  
RYLVSGLVAGTGAIGEGDLRWLVLMWESHAPIANNLMNDLYDDVGTDSATYARA  
MTAHPAATGANAAVYTPRTTSCGSPERALPTTPRWARAVGRSCWRSSPTGCCAP  
RC"  
complement(971.1741)  
/gene="MT0489"  
complement(971.1741)  
/gene="MT0489"  
/note="similar to GP:3063879; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, TetR family"  
/protein\_id="AAK44712.1"  
/db\_xref="GI:13879996"  
/translation="WCODPOLAVVSGASSEWELATVAERIPAVITKDGKRRWHQH  
KVERNELVGTIEAIRHGRFSLMSDIAAEIGVSKTVLYRFVKNLDTTAVMWRFT  
OTLLPNKXALSMDGDFELTIELIVVETVAAOPEYRFVWANSASKSVIADS  
ERITARMALVNRMRQAGNDGGVEPWYLVGGVQLATHSMDSPPRSSDELIDY  
LTMLSWALCGIVEAGGSLEKFEQPHSPFIVPAWGV"  
1956.3182  
/gene="MT0490"  
1956.3182  
/gene="MT0490"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK44713.1"  
/db\_xref="GI:13879997"  
/translation="WALRMKVAVGVFLGATGVFLACRAQADGADHAWLYGMAA  
EAGVYGALADWFAVTFALFKPLGPIPIHTALIKKKDGLGTFVRENFLSPVVE  
TKLDAQIPSLRSLSEATHAQVAETAIVLVVELLRDEIQOVIDRMIVRRIA  
EPQWGPAGRVLATLLAENROEAFIQLLADRAFQWNLGVVIRQVREPSWSPF  
IDHLVGRIRHELMFTDKVRNPDHELRSATFLDFDADLQHDPAIVARDAIKE  
ELMARDSTIAAAAKTKLKLVLGVDDPSALRTITDAVIRIGESLRDADLRDK  
VDSVTVRAAQHLVSEYGVETAITITETIERWDAEASRRRIELHVGRDLQIRINGTV  
GAVAGLAIVAYAAQLF"  
3269.3691  
/gene="MT0491"  
3269.3691  
/gene="MT0491"  
/note="similar to GB:X66407 PID:46399; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, Pbx family"  
/protein\_id="AAK44714.1"  
/db\_xref="GI:13879998"  
/translation="WSSSEKLAKVSTKASVDIGSFIRSQRETAHVSVRQLAERS  
GVSNPYLSQVERGLRKPASADVLSQLAKALRVSAEVLVVRAGILEPSETSQRDAIITD  
TAITERQKILLIDIYASPTHQNEATRECPSPDPTDD"  
complement(3730.3867)  
/gene="MT0492"  
complement(3730.3867)  
/gene="MT0492"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44715.1"  
/db\_xref="GI:13879999"  
/translation="MSGPADPVTTCLCTSTESLLASAADRLDQPAVDLTQMTVLQ  
V"  
4030.4644  
/gene="MT0493"  
4030.4644  
/gene="MT0493"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK44716.1"

gene  
CDS  
/db\_xref="GI:13880000"  
/translation="MKGIPMAENSNDIDKAPLLAALGAADLALATVNELITNLERRA  
EETRDTSRVRESRLTKIQEDLPQLTELREKFTABELRKAAGLYLEAATSYNE  
LVERGAALERLSQSQSEFVSARAXEYVDQAVELTQALGTVAQSTRAVGERAAKLV  
GTELPKAAAPAKKAAAPAKKAAKAAKAPAKKAAAKVTKQ"  
complement(4657.4809)  
/gene="MT0494"  
complement(4657.4809)  
/gene="MT0494"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44717.1"  
/db\_xref="GI:13880001"  
/translation="MHRVHRCHQDQDOYQHDGTHEVAHAFRLCGLHPTRLRLVAPV  
ATRSRW"  
4925.5470  
/gene="MT0495"  
4925.5470  
/gene="MT0495"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44718.1"  
/db\_xref="GI:13880002"  
/translation="MCSMRCPVPVPACIWSCTGSPFSRPRASRANGKALVAVSVA  
VALLGVSAQADPADPGAGBANYGGPSPRLVDHTHWAQWGLPSLRVYPSQVGR  
TASRLGMAADAAWAEVLALSPEADTAGMRAQFICHWQYAEIROFGKFSWNLEWRP  
VVDDEMLASGDCNPGSPESF"  
5470.6144  
/gene="MT0496"  
5470.6144  
/gene="MT0496"  
/note="similar to SP:P39121 PID:809660 PID:1408506  
GB:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="deoxyribose-phosphate aldolase"  
/protein\_id="AAK44719.1"  
/db\_xref="GI:13880003"  
/translation="MLGQPTRAQLAALVDHTLLKPTTADVAALVAEAEIGVYAVC  
VSPSVVAVQAGGVAAVTFPGSKGVSSVKAHEAAALAGASEIDMVIDIGAL  
CGDIDAVRSDIEAVRAAAAGAVLKVIVESAIVLIGQSNATLVDCAAEDAGADFVK  
STGCHPAGGATVRAVELMAETVGPRLGVKASGIRTAADAVAMLAGATRLGLSGTRA  
VLDGLS"  
complement(6169.7215)  
/gene="MT0497"  
complement(6169.7215)  
/gene="MT0497"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK44720.1"  
/db\_xref="GI:13880004"  
/translation="MTNPQGPNDPSFWARPQDGLPAPASSESTQRLRGPBPAG  
HIOEPVSPPTQPPQPTQTEHLAAASHATRSRQAQAQWDPTGLLAAQEEEPAAVK  
KREARDELTVFLVLIIVSLVLAGLIGGELVYHANGVAKVAVCVVQDQATASFG  
VAPLLMQVATHEFTNISVETAGNQRDAKGMQIKLTIONVRLKPNSEGTIGALDA  
TIWSEGIKESQNAIPILGARVTSVTHPADGTVELKGLLNITAKPIVAGKGLE  
LQIINFTNGFSLPKPTVSTLNFTSSLTNPKYPLGIHADSVQVTSVGVSRPSTDA  
AIPFTGQNPFCFSH"  
complement(7212.8234)  
/gene="MT0498"  
complement(7212.8234)  
/gene="MT0498"

Query Match 100.0%; Score 26; DB 1; Length 15670;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGAGCCGACTACTCTGGGT 26  
 |||||  
 Db 4658 GATTCGAGCCGACTACTCTGGGT 4633

RESULT 4  
 AX704277/c 86114 bp DNA linear PAT 03-APR-2003  
 LOCUS  
 DEFINITION  
 AX704277 Sequence 648 from Patent WO02074903.  
 AX704277  
 VERSION  
 AX704277.1 GI:29538532  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.

REFERENCE  
 1  
 Cole, S.  
 Comparative mycobacterial genomics as a tool for identifying  
 targets for the diagnosis, prophylaxis or treatment of  
 mycobacterioses  
 Patent: WO 02074903-A 648 26-SEP-2002;  
 INSTITUT PASTEUR (FR)

FEATURES  
 Location/Qualifiers  
 1..86114  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1773"  
 /note="BAC-Rv265 deposited at the C.N.C.M. under the A.N.  
 I-2628"

ORIGIN  
 Query Match 100.0%; Score 26; DB 6; Length 86114;  
 Best Local Similarity 100.0%; Pred.No. 0.062;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGAGCCGACTACTCTGGGT 26  
 |||||  
 Db 53007 GATTCGAGCCGACTACTCTGGGT 52982

RESULT 5  
 BX248335/c 324050 bp DNA linear BCT 11-JUN-2003  
 LOCUS  
 DEFINITION  
 BX248335 Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment  
 2/14  
 ACCESSION  
 BX248335 BX248333  
 VERSION  
 BX248335.1 GI:31617046  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium bovis subsp. bovis AF2122/97  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.

REFERENCE  
 1  
 Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,  
 Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,  
 Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,  
 Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. and  
 Hewinson, G.  
 The complete genome sequence of Mycobacterium bovis  
 Online Publication  
 PNAS 10.1073/pnas.1130426100 ( Microbiology )  
 2 (bases 1 to 324050)  
 REFERENCE  
 2  
 Garnier, T.  
 Direct Submission  
 Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
 Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the  
 Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
 Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
 Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,

FEATURES  
 Location/Qualifiers  
 1..324050  
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"  
 /mol\_type="genomic DNA"  
 /strain="AF2122/97"  
 /db\_xref="taxon:233413"  
 132..2027  
 /locus\_tag="Mb0290"  
 132..2027  
 /locus\_tag="Mb0290"  
 /notes="Mb0290, -, len: 631 aa. Equivalent to Rv0282, len:  
 631 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (99.8% identity in 631 aa overlap). Conserved hypothetical  
 protein, similar to Y14967/MCB628.18c hypothetical  
 protein from Mycobacterium leprae (573 aa), FASTA scores:  
 opt: 916, E(): 0, (38.7% identity in 568 aa overlap). Also  
 similar to Mycobacterium tuberculosis proteins e.g.  
 Z94121|MTY15F10.26 (619 aa), FASTA scores: opt: 743, E():  
 0, (29.9% identity in 612 aa overlap). Member of CFQX,  
 CBXP family - 9 members in Mycobacterium tuberculosis.  
 Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="CONSERVED HYPOTHETICAL PROTEIN"  
 /protein\_id="CAD93154.1"  
 /db\_xref="GI:31617047"  
 /translation="MAGVGAGSGGVERDDICNVAAASPVASRVNKVDADVVRPAC  
 CRAAGIYQRRPPDLAARGFAALTRVHADQCDQDWTGLAAAGDQSGIVLEAASRT  
 ATTAVLQROVELADNALGLDYTLGLYLFRTATPDGDFHLAYALALSTGPEEFAKT  
 NHVSGITERRAGWRAARLVINVRABSDVVKLLTPWVNDPLDEAFSHAAKIT  
 LGTLARLGMFAPALSYLEPDGPVAAVADGALALVLRHAFVDEEASFEVLOLYA  
 AHPEQVBOALSDDTSFGVITTTAGRIEARTDPDPAPEAGAEVDPAAHERKALL  
 HEALQLAEFIGLDEKRVSELKSSVAMELVKRGSLTVAQRTHLVLPACPGTGT  
 TIARVVKYICGLLLKREINREVRADLIGHIGETAKTNAIDSLDGLVDFDEA  
 YALVATGAKNDLGLVAIDTLARMENDRLVLIAGYADLDKFDLTNEGRLSRFL  
 NIDFYSYTHSELVEIAHKMAEQDSVFQSAHLDEALFALAEASTPTDNGLSRSL  
 DIAGNRFVRNIVERSEERFLDHSAGSGSEFDEELMTITADVVGRSVEPLLRG  
 LGLSVRA"

2024..3640  
 /locus\_tag="Mb0291"  
 2024..3640  
 /locus\_tag="Mb0291"  
 /note="Mb0291, -, len: 538 aa. Equivalent to Rv0283, len:  
 538 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (100.0% identity in 538 aa overlap). Possible conserved  
 membrane protein, similar to several hypothetical  
 mycobacterial proteins e.g. Z94121|MTY15F10.16|RV3895C  
 from Mycobacterium tuberculosis (495 aa), FASTA scores:  
 opt: 698, E(): 0, (37.6% identity in 492 aa overlap);  
 RV1782; RV3450c; RV3869; and Y14967/MCB628.16|MCB628.17c  
 from Mycobacterium leprae (481 aa), FASTA scores: opt:  
 672, E(): 1.5e-31, (37.2% identity in 506 aa overlap).  
 Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"  
 /protein\_id="CAD93155.1"  
 /db\_xref="GI:31617048"  
 /translation="MTNQOHDDHDFDRRSFASRTPVNNPDKVYRFGFTRHQVGT  
 WFFVWRTAAGIALHDTMLVDPRTQSAVLVGLVITGLIGSFVSLRPRGQAG  
 SNAVLADSTALYVRVSEQLHPVNLNLSARLIVGRVPSPTIVKSTELDQFPRNLIG  
 IPGASRWQNTSTDANNVTCVGLNAPSRGAGDGVGVTVIAGPLEDGTARAAALGPGL  
 AYLVDGQETLLWDGKSPIDLADHVTGLGGLGVDPAPRIATAGLFAIPAPPL  
 TAPITPDAGNPAFGVPAPVAVSSYALKDSKGTISDTQVYAVLPGLQOISPLVA  
 ATLIRNNNGYGGQPPRLGADVAKLPSVRLDTRYPSEVYLVVDVTPVTCAYWSK  
 PVGATSLTLTLAGSALPVPDAVHTVELVAGNGGVAIVLAAGTGYFTQVGGPDP  
 APGASLFWSDTGYRVCIDNEPQGVAGGKAVEALGILNPPPIPWNSVLSLFPVGPPT



**TITLE** Direct Submission  
**JOURNAL** Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
**COMMENT** On or before Nov 15, 2003 this sequence version replaced  
 gi:3261524, gi:3261526, gi:3261527, gi:3261529, gi:3242298,  
 gi:3261506, gi:3261689, gi:3261708, gi:3261703, gi:3261760,  
 gi:3261781, gi:3261797, gi:3261800, gi:3261825, gi:3261837.

**Notes:**  
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)).

# FEATURES

## source

1. .342416  
 /organism="Mycobacterium tuberculosis H37Rv"  
 /mol\_type="genomic DNA"  
 /strain="H37Rv"  
 /db\_xref="taxon:83332"  
 223..2118  
 /locus\_tag="Rv0282"  
 223..2118  
 /locus\_tag="Rv0282"  
 /function="UNKNOWN"

## gene

/note="Rv0282, (MTV035.10), len: 631 aa. Conserved hypothetical protein, similar to Y14967|MCB628.18c hypothetical protein from Mycobacterium leprae (573 aa), FASTA scores: opt: 916, E(): 0, (38.7% identity in 568 aa overlap). Also similar to Mycobacterium tuberculosis proteins e.g. 294121|MTY15F10.26 (619 aa), FASTA scores: opt: 743, E(): 0, (29.9% identity in 612 aa overlap). Member of CFXQ, CXP family - 9 members in Mycobacterium tuberculosis. Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental

## CDS

/transl\_table=11  
 /product="CONSERVED HYPOTHETICAL PROTEIN"  
 /protein\_id="CAAL7357.1"  
 /db\_xref="GI:2909471"  
 /translation="NAGVGGDGGVRRDDIGWAASPVASRVNGKVDADVGRFATC  
 CRALGVYQKRPDLAARSGFALTRVAHDQCDANTGLAAGDSIGVLEAASRT  
 ATAGVQVRLADNALGFLDTGLIFRFRATGDDPHLALAAALSTGGPEFAKAT  
 NHVSGITERRAGWAALWLVINRAERSDVVKLLTPMNDPDLDEAFSHAOKIT  
 LGTALRGFMFAPALSYLSEEDPGPVAAVADGAKALVLAHVDESSAELQOLYA  
 APNEQVEALSDDTSGFIVTTAGRIEARDPMDPAPEGAEDFVDPAAHERKAALL  
 HEALQLAEFIGLDEVKQVRLKSSVAMELVKRGITVAQRTHLVAFAGPCTGKT  
 TIARVAKYCGLLGKREINREVRADLIGHIGETEAKTNALIDSALDGLVLEDA  
 VALVATGAKDFGLVAIDTLARMENDRLVLIAGVRADLFLDNEGLRFRTR  
 NIDFSYSHVELVIAHWKARQDSVFQSAHLDEALFALAAESTPDITDNGLSRRSL  
 DIAGNRFVRNIVERSEEREFRLDHSEAGSGFSDBELMTITADDDVGRSVEPLRG  
 LGLSVRA"  
 1375..1398  
 /locus\_tag="Rv0282"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 2115..3731  
 /locus\_tag="Rv0283"  
 2115..3731  
 /locus\_tag="Rv0283"  
 /function="UNKNOWN"

## misc\_feature

/note="Rv0283, (MTV035.11), len: 538 aa. Possible conserved membrane protein, similar to several hypothetical mycobacterial proteins e.g.  
 294121|MTY15F10.16|RV3895c from Mycobacterium tuberculosis (495 aa), FASTA scores: opt: 698, E(): 0, (37.6% identity in 492 aa overlap); Rv1782; Rv3450c; Rv3869; and Y14967|MCB628.16|MCB628.17c from Mycobacterium leprae (481 aa), FASTA scores: opt: 672, E(): 1.5e-31, (37.2% identity in 506 aa overlap). Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental

## gene

/transl\_table=11  
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"  
 /protein\_id="CAAL7358.1"  
 /db\_xref="GI:2909472"  
 /translation="MTNQOHDHDFDHRRSFASRTVNNNPDKVYVRGVTTHQVTVG  
 WRFVRRIRAGIALHDTMLVLDLRTQSRVAVLGLVITGLIGSVFSLIRPNGOAG  
 SNAVLADRSTAAALVYRGEQHLVNLTSRLVIGRVSPTTVKSTELDQFPRGNLIG  
 IPGERWQNTSTNDANVTCDGNAPSGAGVGVTIAGIELDGTARAAALPGQG  
 AVLDGAGTMLWDKGRSPIDLAHVATSGLGADVPAPRIIASGLFNAPAEPL  
 TATIPDAGNFWASFGVPAPICAVSSVALDKGKTSIDTVQYVAVLDGLQIISPLA  
 AILNNNSYGLQOPRLGADAVAKLPVSRVLDTRRYPSPVSLVDTRDPTVCAVSK  
 PVGATSSLLAGSLPVEDAVLVELVAGNGGVAATVALAAGNGYFTOTVGGGPD  
 APGAGSLFWSDTGVRYGIDNEPQGVAGGKAVEALGLNPPVPPIPKSVLSLFPVPT  
 LSRDALLADHTLVPSRPARPVSAEGGR"

## CDS

/locus\_tag="Rv0284"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 3728..7720  
 /locus\_tag="Rv0284"  
 3728..7720  
 /locus\_tag="Rv0284"  
 /function="UNKNOWN"  
 /note="Rv0284, (MTV035.12), len: 1330 aa. Possible conserved membrane protein, similar to products of two adjacent Mycobacterium leprae genes, MLCB628.16c (744 aa) and MLCB628.15c (597 aa); and throughout its length to several large Mycobacterium tuberculosis proteins: Rv3447c, Rv3870, Rv1784, etc. Y14967|MCB628.15 (744 aa), FASTA scores: opt: 942, E(): 0, (33.8% identity in 730 aa overlap); Y14967|MCB628.14 (597 aa), FASTA scores: opt: 613, E(): 3.1e-30, (31.7% identity in 615 aa overlap); 294121|MTY15F10.17 (1396 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap); 295389|MCY77.19 (1236 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap). Contains PS00017 ATP/GTP-binding site motif A (P-loop)."  
 /codon\_start=1  
 /evidence=experimental

## misc\_feature

/locus\_tag="Rv0284"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 6212..6235  
 /locus\_tag="Rv0284"

## misc\_feature

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 7046..7069

## misc\_feature

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 7046..7069

/locus tag="Rv0284"  
 /note="FS00017 ATP/GTP-binding site motif A (P-loop)"  
 7717..8025  
 /gene="PE5"  
 /locus tag="Rv0285"  
 7717..8025  
 /gene="PE5"  
 /locus tag="Rv0285"  
 /function="UNKNOWN"  
 /note="Rv0285, (MTV035.13), len: 102 aa. Member of the  
 Mycobacterium tuberculosis PE family (see Brennan & Delogu  
 2002), similar to others e.g. AL0212|MTV012.37 from  
 Mycobacterium tuberculosis (105 aa), FASTA scores: opt:  
 497, E(): 2.6e-24, (80.4% identity in 102 aa overlap);  
 280108|MTV02184.03 from Mycobacterium tuberculosis (102  
 aa), FASTA scores: opt: 413, E(): 3.7e-19, (66.7% identity  
 in 102 aa overlap); etc."  
 /codon\_start=1

Query Match 100.0%; Score 26; DB 1; Length 342416;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GATTCGAGCCGACTACTCTCGGT 26  
 |||||  
 Db 224503 GATTCGAGCCGACTACTCTCGGT 224478

RESULT 7  
 BX465204  
 LOCUS  
 DEFINITION  
 BX465204 168776 bp DNA linear VRT 04-JUL-2003  
 Zebrafish DNA sequence from clone CH211-160N19 in linkage group 22,  
 complete sequence.  
 ACCESSION  
 BX465204.5 GI:32452245  
 VERSION  
 HTG.  
 KEYWORDS  
 Danio rerio (zebrafish)  
 ORGANISM  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 168776)  
 Barlow,K.  
 Direct Submission  
 Submitted (04-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk  
 On Jul 4, 2003 this sequence version replaced gi:30962381.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information  
 on the WormPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Zebrafish pUC subclones occasionally display inconsistency over the  
 length of mononucleotide A/T runs and conserved TA repeats. Where  
 this is found the longest good quality representation will be  
 submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat  
 discovery system (Zhirong Bao and Sean Eddy, submitted); and those  
 beginning 'drr' were identified by Rick Waterman (Stephen Johnson  
 lab, WashU). For further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml)  
 CH211-160N19 is from a CHORI-211 BAC library

VECTOR: PTARBAC2.1.  
 Location/Qualifiers  
 1..168776  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="CH211-160N19"  
 /clone\_lib="CHORI-211"

## ORIGIN

Query Match 75.4%; Score 19.6; DB 5; Length 168776;  
 Best Local Similarity 84.6%; Pred. No. 1.1e+02;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTCGAGCCGACTACTCTCGGT 26  
 |||||  
 Db 62685 GTTACGAGCGCGACTACTCTCGGT 62710

## RESULT 8

HSDJ723E3/c  
 LOCUS  
 DEFINITION  
 HSDJ723E3 73112 bp DNA linear PRI 11-MAR-2001  
 Human DNA sequence from clone RP4-723E3 on chromosome 20 Contains  
 ESTs, STSs, and GSSs, complete sequence.  
 ACCESSION  
 AL049736  
 VERSION  
 AL049736.10 GI:5531257  
 KEYWORDS  
 HTG.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 73112)  
 Skuce,C.  
 Direct Submission  
 Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Jul 19, 1999 this sequence version replaced gi:5441417.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em:, EMBL; Sw:.,  
 SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information on the WormPEP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 20, constructed by the Sanger Centre Chromosome 20  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>  
 RP4-723E3 is from the library RPCI-4 constructed by the group of  
 Peter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>



## VECTOR: PCVPAC2

IMPORTANT: This sequence is not the entire insert of clone RP4-723E3 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP4-723E3 is at 1 in this sequence. The true left end of clone RP4-715N11 is at 73013 in this sequence. The true right end of clone RP5-1022J11 is at 33305 in this sequence.

## FEATURES

```

source
    Location/Qualifiers
        1..73112
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP4-723E3"
            /clone_lib="RPCI-4"
        2..634
            /notes="match: GSS: Em:AQ478861"
        33..461
            /notes="match: STS: Em:AL022263"
        750..934
            /note="MER20 repeat: matches 20..207 of consensus"
        1257..1465
            /note="MIR repeat: matches 15..232 of consensus"
        2187..2664
            /note="match: GSS: Em:AQ499959"
        2703..2753
            /note="MIR repeat: matches 158..209 of consensus"
        2967..3178
            /note="MER20 repeat: matches 1..218 of consensus"
        3211..3624
            /note="MER69B repeat: matches 3..524 of consensus"
        3625..3929
            /note="AluSx repeat: matches 1..309 of consensus"
        3930..4053
            /note="MER69B repeat: matches 524..649 of consensus"
        4054..4249
            /note="AluA13 repeat: matches 5948..6143 of consensus"
        4250..4810
            /note="MER69B repeat: matches 620..1200 of consensus"
        4833..4964
            /note="L2 repeat: matches 1668..1793 of consensus"
        5000..5079
            /note="MLT1A2 repeat: matches 90..170 of consensus"
        5080..5347
            /note="AluY repeat: matches 1..267 of consensus"
        5348..5555
            /note="MLT1A2 repeat: matches 170..371 of consensus"
        5564..5999
            /note="MLT1C repeat: matches 24..466 of consensus"
        6386..6549
            /note="L2 repeat: matches 2301..2475 of consensus"
        7109..7165
            /note="Alu repeat: matches 244..300 of consensus"
        7391..7454
            /note="CRNA-11e-ATT repeat: matches 9..72 of consensus"
        8052..8278
            /note="MIR repeat: matches 8..252 of consensus"
        8335..8418
            /note="L2 copies 2 mer to 84% conserved"
        8419..8466
            /note="L2 copies 2 mer to 100% conserved"
        9162..9353
            /note="MER20 repeat: matches 1..218 of consensus"
        9439..9657
            /note="MIR repeat: matches 42..262 of consensus"
        10679..10833
            /note="MLT1A1 repeat: matches 1..155 of consensus"
        10834..11134
            /note="AluSx repeat: matches 1..297 of consensus"
        11135..11342
            /note="MLT1A1 repeat: matches 155..363 of consensus"
        12318..12626
            /note="AluSc repeat: matches 1..307 of consensus"
repeat_region
    13542..13951
        /note="L1MC1 repeat: matches 5919..6332 of consensus"
        complement(15552..15832)
        /note="match: STS: Em:T99480"
    15833..15925
        /note="MLT1J repeat: matches 227..318 of consensus"
    16330..16713
        /note="L28 copies 3 mer etc 54% conserved"
    16931..17237
        /note="AluJb repeat: matches 5..307 of consensus"
    17239..17633
        /note="match: GSS: Em:AQ531284"
        complement(17550..17937)
        /note="match: GSS: Em:AQ381104"
    17938..18045
        /note="MLT1E repeat: matches 2..111 of consensus"
    18046..18401
        /note="THE1B repeat: matches 1..364 of consensus"
    18402..18874
        /note="MLT1E repeat: matches 111..568 of consensus"
    18962..19037
        /note="L38 copies 2 mer ag 64% conserved"
    20108..20189
        /note="L41 copies 2 mer cc 73% conserved"
    20460..20585
        /note="MLT1H repeat: matches 373..503 of consensus"
    20739..20924
        /note="MIR repeat: matches 2..206 of consensus"
    22955..23253
        /note="AluY repeat: matches 1..298 of consensus"
    25295..25624
        /note="L2 repeat: matches 2012..2319 of consensus"
    25625..26045
        /note="MLT1C repeat: matches 3..466 of consensus"
    26046..26377
        /note="L2 repeat: matches 2319..2702 of consensus"
    26481..26559
        /note="L2 repeat: matches 2651..2737 of consensus"
    26592..26710
        /note="MIR repeat: matches 123..250 of consensus"
    28531..28600
        /note="L35 copies 2 mer ta 65% conserved"
    28641..28697
        /note="LIP3 repeat: matches 5600..5656 of consensus"
    28978..29151
        /note="MIR repeat: matches 105..256 of consensus"
    29159..29281
        /note="AluJ repeat: matches 6..127 of consensus"
    29309..29451
        /note="AluJ repeat: matches 118..258 of consensus"
    29869..30055
        /note="MLT1A2 repeat: matches 180..366 of consensus"
    30051..30138
        /note="MLT1A1 repeat: matches 79..172 of consensus"
    30164..30458
        /note="AluSg repeat: matches 1..295 of consensus"
    30487..30545
        /note="MLT1A2 repeat: matches 68..126 of consensus"
    30546..30585
        /note="VADE1 repeat: matches 40..80 of consensus"
    30598..30622
        /note="MLT1F repeat: matches 355..379 of consensus"
    30646..30695
        /note="MLT1F repeat: matches 2..51 of consensus"
    30696..30759
        /note="MLT1A2 repeat: matches 5..68 of consensus"
    30945..31239
        /note="AluSx repeat: matches 1..295 of consensus"
    31243..31549
        /note="AluSc repeat: matches 1..309 of consensus"
    31603..32075
        /note="MLT1F repeat: matches 61..537 of consensus"
    32076..32374

```



```

/note="AluSc repeat: matches 1. .299 of consensus"
32491. .32611
/note="L1 repeat: matches 2617. .2748 of consensus"
32984. .33352
/note="L1P repeat: matches 3840. .4219 of consensus"
33353. .33829
/note="L1PA7 repeat: matches 5666. .6141 of consensus"
complement(33830. .34289)
/note="match: GSS: Em:B53575"
complement(33830. .34282)
/note="match: GSS: Em:AQ224090"
34561. .34636

Query Match      73.1%; Score 19; DB 9; Length 73112;
Best Local Similarity 100.0%; Pred.No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GGAGCCCGACTACTCTCG 24
DB 70586 GGAGCCCGACTACTCTCG 70568

RESULT 9
AC072036/c
LOCUS AC072036
DEFINITION Homo sapiens chromosome 3 clone RP11-269016, WORKING DRAFT
ACCESSION AC072036
VERSION AC072036.6 GI:20335678
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 150881)
Muzny,D.M., Adams,C., Adio-osman,D., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burrett,C., Surrrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoggess,M., Holloway,C., Hollins,B.,
Homai,F., Howard,S., Huber,U., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozaco,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshcari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Tansley,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Woodden,S., Worley,K.,

```

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 150881)  
 Worley,K.C.  
 Direct Submission  
 Submitted (07-JUN-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Apr 28, 2002 this sequence version replaced gi:1849599.  
 -----  
 Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: HBMF  
 Center clone name: RP11-269016  
 -----  
 Summary Statistics  
 Sequencing vector: M13;  
 Chemistry: Dye-primer Bodipy: 19% of reads  
 Chemistry: Dye-terminator Big Dye: 81% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 146144 bases at least Q40  
 Consensus quality: 150709 bases at least Q30  
 Consensus quality: 153535 bases at least Q20  
 Estimated insert size: 149311; sum-of-contigs estimation  
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 2185: contig of 2185 bp in length  
 \* 2186 2285: gap of unknown length  
 \* 2286 5263: contig of 2978 bp in length  
 \* 5264 5363: gap of unknown length  
 \* 5364 8405: contig of 3042 bp in length  
 \* 8406 8505: gap of unknown length  
 \* 8506 13401: contig of 4896 bp in length  
 \* 13402 13501: gap of unknown length  
 \* 13502 13297: contig of 5796 bp in length  
 \* 13298 13997: gap of unknown length  
 \* 13998 22990: contig of 10593 bp in length  
 \* 22991 30090: gap of unknown length  
 \* 30091 39320: contig of 9230 bp in length  
 \* 39321 39420: gap of unknown length  
 \* 39421 52899: contig of 13479 bp in length  
 \* 52900 52999: gap of unknown length  
 \* 53000 75579: contig of 22579 bp in length  
 \* 75579 95768: gap of unknown length  
 \* 95769 99888: contig of 24110 bp in length  
 \* 99889 124551: contig of 24663 bp in length  
 \* 124552 150881: gap of unknown length  
 \* 150882 150881: contig of 26230 bp in length.  
 Location/Qualifiers  
 1..150881  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"

ORIGIN  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-269016"

Query Match 73.1%; Score 19; DB 2; Length 150881;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GGAGCCCGACTACTTCTGG 24

Db 62195 GGAGCCCGACTACTTCTGG 62177

RESULT 10  
 BC028151/c  
 LOCUS  
 DEFINITION  
 Homo sapiens DNA segment on chromosome X and Y (unique) 155  
 expressed sequence, mRNA (cDNA clone MGC:39904 IMAGE:5217247),  
 complete cds.

ACCESSION  
 BC028151 GI:20380085  
 VERSION  
 MGC.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)

REFERENCE  
 AUTHORS  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2187)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Lequellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worthy, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257

2 (bases 1 to 2187)  
 Strausberg, R.

Direct Submission  
 Submitted (08-APR-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@nigr.nih.gov](mailto:nisc_mgc@nigr.nih.gov)

Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granat, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,

Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W.,  
 Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 62 Row: j Column: 19  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 10835221.

FEATURES  
 source

1..2187  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:39904 IMAGE:5217247"  
 /tissue\_type="Blood, adult leukocytes"  
 /clone\_id="NIH MGC\_118"  
 /lab\_host="DH10B"  
 /notes="vector: pCMV-SPORT6"

gene

1..2187  
 /gene="DXYS155E"  
 /notes="synonyms: XE7, XE7Y"  
 /db\_xref="LocusID:8227"  
 /db\_xref="MIM:465000"  
 165..1505  
 /codon\_start=1  
 /product="DXYS155E protein"  
 /protein\_id="AAH28151.1"  
 /db\_xref="GI:20380086"  
 /db\_xref="LocusID:8227"

CDS

translation="MAAAIVHTSEAVELCPAYGLYLPKMTISVALPOLKQPK  
 STSNVMEKMGVQHQFSTLRISKSTMDFFRFEVGEVSEVSLVSLACLDGKTKT  
 LKGSFDILKVAAEFKIDFTDHDVDFRDADKNETLPGRPDTTHLEGLPCKWFA  
 LKSGSEKFSVDVVKFKFGEIRNVDIPMDSYRENTGFRNFTTFGGHNFAY  
 VQRYNGFTQAMSLRGMKLMYKGDGKAVACNIKVSFDTKLSADAIKKQRLERQ  
 KLCLQEOREOKREKAEQRAERQKLEERERERERERERERERERERERERER  
 RRNKKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK  
 RQRPQCPQCPPLKCGRRGAVSPAAAVATKPAALMPRTAFSREGVALVCSR"

ORIGIN

Query Match 72.3%; Score 18.8; DB 9; Length 2187;  
 Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGAGCCCGACTACTTCTGGG 25  
 Db 664 TCGAGCCCGACTACTTCTGGG 643

RESULT 11  
 HUMLSPRO/c

LOCUS  
 DEFINITION  
 Human lymphocyte surface protein exons 1-5, complete cds.  
 ACCESSION  
 M99578

VERSION  
 M99578.1 GI:187241  
 KEYWORDS  
 Lymphocyte surface protein.  
 SOURCE  
 Homo sapiens (human)

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 Voland, J.R., Wzykowski, R.J., Huang, M. and Dutton, R.W.  
 Cloning and sequencing of a trophoblast-endothelial-activated  
 lymphocyte surface protein: cDNA sequence and genomic structure

Proc. Natl. Acad. Sci. U.S.A. 89 (21), 10425-10429 (1992)  
 93066251

JOURNAL  
 MEDLINE  
 PUBMED  
 1438229

COMMENT  
 Original source text: Homo sapiens (library: lambda gt11) neonate  
 placenta cDNA to mRNA.  
 Placenta Location/Qualifiers  
 1..2622  
 /organism="Homo sapiens"

```
/mol_type="mrna"
/db_xref="taxon:9606"
/cell_type="trophoblast"
/tissue_type="placenta"
/dev_stages="neonate"
/tissue_lib="ncbi"
/tissue_lib="lambda gt11"
join(1..153,153..933,934..1082,1083..1323,1324..2622)
1..171
/note="84% GC in 5'UTR"
1..153
/note="codes the 5' UTR; intron between exon 1 and exon 2
is 2.7Kb"
/number=1
153..933
/note="codes the extracellular domain; intron between exon
2 and exon 3 is 0.7Kb"
/number=2
170..1822
/note="550 amino acids MW=61kDa, glycosylated=75 kDa;
expressed on endothelium, activated lymphocytes and
syncytiotrophoblast, contains leucine zipper and basic
region homologous to myc; 721p"
/codon_start=1
/protein_id="AAA36187.1"
/db_xref="GI:187242"
/translation="MAAATIVHDTSEAVELCPAYGLYLKPTKMTISVALPOLKQPKG
SISNWEVRLKGMVQNHQFSTLRISKSTMDIFRFEVEVENKSLVKSFLACLQKTK
LSGFSDLIKVRAAEFKIDFTRHDWDSFFRDAKDMNETLGERPDTIHLEGLPCKWFA
IKSSGSEKPSDEVLVKVFKEFGRINVDIPMLDPYREEMTGRNFTSFSGHLNFEAY
VQREYMGFTIQMSALRGKMLYKGDGKAVACNIKVSFDTKHLSDASIKKQLEQ
KLQELQOEQBEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
RRNQKLEKQAEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
OKEEKLQOEQBEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
HTHDLGVAGHADLQVLDILOTVSSGVSATLHPLGGOPAGAPKESAAHPDGA
PKSVNGSVAEAPCKEQVSSCRVVPEDGSEKPCGGLSCIENNQQKGIACEON
VSKRDTSEODKCNRFSGKGRATGDLADRRKRSRRARRASSEDGPRKRRRH
KQAYKDDSPNRRSTPDHTRSRSHSKDRHRSRRRRSGASRKHSHRRRSER
SRSPSRHRSSTNR"
172..232
/note="codes for protein leader sequence"
934..1082
/note="codes the putative basic region; intron between
exon 3 and exon 4 is 3.5Kb"
/number=3
1083..1323
/note="codes the leucine zipper region; the intron between
exon 4 and exon 5 is 1.3Kb"
/number=4
1199..1263
/note="codes for leucine zipper, homologous to myc"
1324..2622
/note="codes the serine phosphorylation site and the
3'UTR"
/number=5

misc_feature
172..232
/note="codes for protein leader sequence"
934..1082
/note="codes the putative basic region; intron between
exon 3 and exon 4 is 3.5Kb"
/number=3
1083..1323
/note="codes the leucine zipper region; the intron between
exon 4 and exon 5 is 1.3Kb"
/number=4
1199..1263
/note="codes for leucine zipper, homologous to myc"
1324..2622
/note="codes the serine phosphorylation site and the
3'UTR"
/number=5

ORIGIN
Query Match 72.3%; Score 18.8; DB 9; Length 2622;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGAGCCCGGACTACTCTGGG 25
Db 669 TCGAGCCCGGACTCTCTCAGG 648

RESULT 12
HUXKFA/c
LOCUS
DEFINITION Human XE7 mRNA, complete alternate coding regions.
ACCESSION L03426
VERSION L03426.1 GI:340386
KEYWORDS XE7 gene; alternative splicing; pseudoautosomal gene.
SOURCE Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
1 (bases 1 to 3233)
Ellison J.W., Ramos, C., Yen, P.H. and Shapiro, L.J.
Structure and expression of the human pseudoautosomal gene XE7
Hum. Mol. Genet. 1 (9), 691-696 (1992)
JOURNAL
MEDLINE
PUBMED
93258310
COMMENT
Original source text: Homo sapiens (tissue library: gt10 cDNA) cDNA
to mRNA.
FEATURES
Location/Qualifiers
1..3233
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/map="X and Y chromosomes"
/cell_line="IMR90"
/tissue_lib="gt10 cDNA"
1..147
/genes="XE7"
/number=1
148..928
/genes="XE7"
/number=2
join(167..928,929..1077,1078..1318,1389..2324)
/genes="XE7"
/note="695 aa isoform"
/codon_start=1
/protein_id="AAA61304.1"
/db_xref="GI:340388"
/translation="MAAATIVHDTSEAVELCPAYGLYLKPTKMTISVALPOLKQPKG
SISNWEVRLKGMVQNHQFSTLRISKSTMDIFRFEVEVENKSLVKSFLACLQKTK
LSGFSDLIKVRAAEFKIDFTRHDWDSFFRDAKDMNETLGERPDTIHLEGLPCKWFA
IKSSGSEKPSDEVLVKVFKEFGRINVDIPMLDPYREEMTGRNFTSFSGHLNFEAY
VQREYMGFTIQMSALRGKMLYKGDGKAVACNIKVSFDTKHLSDASIKKQLEQ
KLQELQOEQBEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
RRNQKLEKQAEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
OKEEKLQOEQBEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
HTHDLGVAGHADLQVLDILOTVSSGVSATLHPLGGOPAGAPKESAAHPDGA
PKSVNGSVAEAPCKEQVSSCRVVPEDGSEKPCGGLSCIENNQQKGIACEON
VSKRDTSEODKCNRFSGKGRATGDLADRRKRSRRARRASSEDGPRKRRRH
KQAYKDDSPNRRSTPDHTRSRSHSKDRHRSRRRRSGASRKHSHRRRSER
SRSPSRHRSSTNR"
join(167..928,929..1077,1078..1318,1319..1324)
/genes="XE7"
join(167..928,929..1077,1078..1318,1319..1324)
/genes="XE7"
/note="385 aa isoform"
/codon_start=1
/protein_id="AAA61303.1"
/db_xref="GI:340387"
/translation="MAAATIVHDTSEAVELCPAYGLYLKPTKMTISVALPOLKQPKG
SISNWEVRLKGMVQNHQFSTLRISKSTMDIFRFEVEVENKSLVKSFLACLQKTK
LSGFSDLIKVRAAEFKIDFTRHDWDSFFRDAKDMNETLGERPDTIHLEGLPCKWFA
IKSSGSEKPSDEVLVKVFKEFGRINVDIPMLDPYREEMTGRNFTSFSGHLNFEAY
VQREYMGFTIQMSALRGKMLYKGDGKAVACNIKVSFDTKHLSDASIKKQLEQ
KLQELQOEQBEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
RRNQKLEKQAEQKQLEKLEERKLEKLLAQNLQIRLIELLSRAKAVKLEQ
929..1077
/genes="XE7"
/number=3
1078..1318
/genes="XE7"
/number=4
1319..1388
/genes="XE7"
/note="alternate"
/number=5
1389..3233
/genes="XE7"
/note="alternate"
/number=6
3223..3228
/genes="XE7"
polyA_signal
3223..3228
/genes="XE7"
ORIGIN
```

```

Query Match          72.3%; Score 18.8; DB 9; Length 3233;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 TCGAGCCCGGACTACTCTCTGGG 25
|||||
Db 666 TCGAGCCCGGACTCTCTTCAGGG 645

RESULT 13
AL683875 184927 bp DNA linear HTG 02-APR-2003
LOCUS Homo sapiens chromosome X clone RP11-446K8, 5 unordered pieces.
DEFINITION AL683875
ACCESSION AL683875.25 GI:29498394
VERSION HTG; HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 184927)
JOURNAL Lawlor S.
Direct Submission
Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 2, 2003 this sequence version replaced gi:28971596.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA446K8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 86% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads
Consensus quality: 182378 bases at least Q40
Consensus quality: 18293 bases at least Q30
Consensus quality: 183854 bases at least Q20
Insert size: 184527; sum-of-contigs
Insert size: 179006; 2.7% error; agarose-fp
Quality coverage: 9.16x in Q20 bases; sum-of-contigs Quality
coverage: 10.59x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 49340: contig of 49340 bp in length
* 49341 49440: gap of 100 bp
* 49441 91479: contig of 42039 bp in length
* 91480 91579: gap of 100 bp
* 91580 94862: contig of 3283 bp in length
* 94863 94962: gap of 100 bp
* 94963 119846: contig of 24884 bp in length
* 119847 119947: gap of 100 bp
* 119947 184927: contig of 64981 bp in length.
FEATURES
Location/Qualifiers
1..184927
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-446K8"
/clone_lib="RPCI-11.2"
misc_feature 1..49340

```

```

/note="assembly_fragment:04719
clone_end:T7
vector_side:left"
49441..91479
/note="assembly_fragment:00421"
91580..94862
/note="assembly_fragment:01738"
94963..119846
/note="assembly_fragment:04642"
119947..184927
/note="assembly_fragment:04655
clone_end:SP6
vector_side:right"

ORIGIN
Query Match          72.3%; Score 18.8; DB 2; Length 184927;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 TCGAGCCCGGACTACTCTCTGGG 25
|||||
Db 76777 TCGAGCCCGGACTCTCTTCAGGG 76798

RESULT 14
AL683807/c
LOCUS Human DNA sequence from clone RP13-297E16 on chromosome X, complete
DEFINITION AL683807
ACCESSION AL683807.22 GI:25251452
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 189825)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Direct Submission
Submitted (22-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 24, 2002 this sequence version replaced gi:24474461.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
En:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP13-297E16 is from the library RPCI-13.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least

```

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

## FEATURES

## source

1..189825  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="RP13-297E16"  
/clone\_lib="RPC1-13.2"

## ORIGIN

Query Match 72.3%; Score 18.8; DB 9; Length 189825;

Best Local Similarity 90.9%; Pred. NO. 2.8e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGGAGCCCGACTACTTCTGGG 25

Db 11422 TCGGAGCCCGACTCTCTCAGG 11401

## RESULT 15

AY398405/c

## LOCUS

AY398405 1542 bp mRNA linear VRT 20-OCT-2003

Definition: Danio rerio clone RK063H4C07 autoantigenic hnRNP-associated with lethal yellow (RALY) mRNA, complete cds.

## ACCESSION

AY398405

## VERSION

AY398405.1

## KEYWORDS

GI:37681920

## SOURCE

Danio rerio (zebrafish)

## ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 1542)

Song, H.D., Wu, X.Y., Sun, X.J., Zhou, Y., Liu, T.X., Deng, M.,

Zhang, G.W., Sheng, Y., Chen, Y., Ruan, Z., Jiang, C.L., Fan, H.Y.,

Zou, L.I., Kanki, J.P., Look, A.T. and Chen, Z.

Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue

Unpublished

2 (bases 1 to 1542)

Zhang, G.W., Sun, X.J., Wu, X.Y., Song, H.D., Zhou, Y., Liu, T.X.,

Deng, M., Sheng, Y., Chen, Y., Ruan, Z., Jiang, C.L., Fan, H.Y.,

Zou, L.I., Kanki, J.P., Look, A.T. and Chen, Z.

Direct Submission

Submitted (24-SEP-2003) State Key Lab for Medical Genomics,

Shanghai Institute of Hematology, Ruijin Hospital Affiliated to

Shanghai Second Medical University, 197 Rui Jin Road II, Shanghai

200025, P. R. China

200025, P. R. China

Location/Qualifiers

1..1542

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="RK063H4C07"

/tissue type="kidney marrow"

1..1542

/gene="RALY"

154..985

/gene="RALY"

/note="RNA binding protein"

/codon\_start=1

/product="autoantigenic hnRNP-associated with lethal

yellow"

/protein\_id="AA097838.1"

/db\_xref="GI:37681921"

/translation="MSLKVQTSNITNDPKINSRVFIGNLNTAVKSDVEIFSK

YGRVLGCSVHKGFYQVNERHARGAVIGENRVLAGQTLDDINWAGPKNRPKGLK

RSALYSGDYDFDYRLFEYRVPVPRVAVPVVRRVKSLLP

VKLLTRSAIIPNSVVKHLKSTELQIKSNIDALIGRLDQITEDYCYSTEL

OKAEDLKSEASQDESSESQDLQHSDEVEGDHTEECDDMMNNHISEMDPIIQ"

## ORIGIN

Query Match 71.5%; Score 18.6; DB 5; Length 1542;

Best Local Similarity 84.0%; Pred. NO. 3.1e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTCCGAGCCCGACTACTTCTGGG 25

Db 873 GATTCCGAGCCCGACTCTCTGAG 849

Search completed: April 29, 2004, 06:01:53

Job time : 396.645 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 214.627 Seconds  
(without alignments)  
514.627 Million cell updates/sec

Title: US-10-624-714-8

Perfect score: 26

Sequence: 1 gattcgagccgactactcttgggt 26

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Jan04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002s.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
C 1	26	100.0	1097	2	AAT86594	Aat86594 Mycobacte
C 2	26	100.0	1097	2	AAT76948	Aat76948 Mycobacte
C 3	26	100.0	86114	6	ABX09143	Abx09143 Mycobacte
C 4	26	100.0	110000	4	AAI99682_05	Continuation (6 of
C 5	26	100.0	110000	4	AAI99683_05	Continuation (6 of
C 6	18.2	70.0	493	3	AA82062	Aa82062 N. mening
C 7	18	69.2	21581	9	ADC87234	Adc87234 Human GPC
C 8	17.6	67.7	288	8	ADB10175	Adb10175 Alloiococ
C 9	17.6	67.7	4268	4	ABL08045	Ab108045 Drosophil
C 10	17.6	67.7	5227	4	AAI37093	AAI37093 Human mus
C 11	17.6	67.7	5227	5	ABA14506	Abal14506 Human ner
C 12	17.6	67.7	5227	5	ABA14501	Abal14501 Human ner
C 13	17.6	67.7	5227	5	ABA17614	Abal17614 Human ner
C 14	17.6	67.7	5227	7	ABX60081	Abx60081 cDNA enco
C 15	17.6	67.7	6670	4	ABL08044	Ab108044 Drosophil
C 16	17.6	67.7	110000	8	ADB12064_12	Continuation (13 o
C 17	17.2	66.2	1029	7	AB266701	Ab266701 Orthosomy
C 18	17.2	66.2	2136	9	ADB62464	Abd62464 Human cDN
C 19	17.2	66.2	2729	4	AA59701	Aa59701 Propionib
C 20	17.2	66.2	2729	7	ACF64630	Acf64630 Propionib
C 21	17.2	66.2	4060	6	ABQ81281	Abq81281 Arabidops
C 22	17.2	66.2	4358	4	AA59648	Aa59648 Propionib
C 23	17.2	66.2	4358	7	ACF64577	Acf64577 Propionib

C 24	17.2	66.2	45055	7	ABZ66808	Abz66808 Orthosomy
C 25	17	65.4	325	3	ABX35150	Abx35150 Bovine ES
C 26	17	65.4	369	3	AAD02041	Aad02041 Human DNA
C 27	17	65.4	384	4	AAI01264	AAI01264 Human rep
C 28	17	65.4	384	4	ABL96723	Ab196723 Human tes
C 29	17	65.4	505	7	ABZ17764	Abz17764 S2 subtra
C 30	17	65.4	697	2	AAx90829	Aax90829 Human sec
C 31	17	65.4	697	3	AAZ52481	Aaz52481 Human sec
C 32	17	65.4	1697	2	AAV65665	Aav65665 Signal tr
C 33	17	65.4	1880	4	ABL14341	Ab114341 Drosophil
C 34	17	65.4	2991	4	ABL04826	Ab104826 Drosophil
C 35	17	65.4	3111	4	AAF26660	Aaf26660 Human Sma
C 36	17	65.4	3111	6	ABL66567	Ab166567 Lung canc
C 37	17	65.4	3111	6	ABK64389	Abk64389 Human ben
C 38	17	65.4	3111	7	AAI60335	AAI60335 Human Sma
C 39	17	65.4	3681	7	AAI60336	AAI60336 Mouse Sma
C 40	17	65.4	3695	7	ACC49355	Acc49355 Human SHT
C 41	17	65.4	3753	4	AAI04936	AAI04936 Human rep
C 42	17	65.4	3753	4	ABL97830	Ab197830 Human tes
C 43	17	65.4	4311	7	AAI60337	AAI60337 Rat Smad-
C 44	17	65.4	4614	4	ABL14340	Ab114340 Drosophil
C 45	17	65.4	4995	4	ABL26790	Ab126790 Drosophil

#### ALIGNMENTS

##### RESULT 1

AAT86594/C

ID AAT86594 standard; DNA; 1097 BP.

XX AC AAT86594;

DT 14-OCT-1998 (first entry)

DE Mycobacterial heparin-binding haemagglutinin antigen gene.

KW Surface protein; Mycobacterium bovis; BCG; adhesion; epithelium; ds;  
KW Bacille Calmette-Guerin; Mycobacterium tuberculosis; epithelial cell;  
KW heparin-binding haemagglutinin antigen; PCR; primer; amplification;  
KW probe; hybridisation; chromosome; vaccine; diagnosis; immunassay.  
XX OS Mycobacterium bovis.  
XX FH Key Location/Qualifiers  
CDS 331..930  
FT /\*tag= a  
FT /product= "HBHA"  
ET /note= "heparin-binding haemagglutinin antigen"

XX FR2748749-A1.

XX 21-NOV-1997.

XX 17-MAY-1996; 96FR-00006169.

XX 17-MAY-1996; 96FR-00006169.

XX (INSP ) INST PASTEUR LILLE.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Menozzi F, Loch C;

XX WPI; 1998-021392/03.

XX P-PSDB; AAW43082.

XX Mycobacterial heparin-binding haemagglutinin polypeptide - useful for  
vaccination against and diagnosis of mycobacterial infections.

XX Claim 19; Fig 10; 50pp; French.

XX This sequence represents the gene encoding a surface protein found on  
Mycobacterium bovis BCG (Bacille Calmette-Guerin) or M. tuberculosis,

CC which enables mycobacteria to adhere to host, especially epithelial,  
 CC cells. The protein was isolated by passing a culture of BCG  
 CC microorganisms over a Heparin-Sepharose column and eluting proteins with  
 CC a gradient of 0-0.5 M NaCl. Fractions were separated by gel  
 CC electrophoresis and a 28 kD heparin binding protein purified. N-terminal  
 CC and internal peptide fractions were sequenced and the amino acid  
 CC sequences used to design PCR primers (AA76596-T86599). These amplified a  
 CC 150 bp fragment (AA76595) used as a probe to isolate chromosomal  
 CC fragments containing the gene which encodes a heparin-binding  
 CC haemagglutinin antigen (HBHA). The protein or peptides derived from it  
 CC (e.g. AA76596-T86599), can be used in vaccines against mycobacterial infections  
 CC (particularly M. bovis or M. tuberculosis), or can be used to diagnose  
 CC mycobacterial infection (by immunoassay detection of anti-HBHA  
 CC antibodies)  
 XX  
 SQ Sequence 1097 BP; 236 A; 324 C; 365 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 1097;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGAGCCGACTACTTCTGGGT 26  
 DB 944 GATTCGAGCCGACTACTTCTGGGT 919

## RESULT 2

ID AA76948/c  
 ID AA76948 standard; DNA; 1097 BP.

XX AA76948;  
 AC AA76948;

XX 14-OCT-1998 (first entry)

DE Mycobacterial heparin-binding haemagglutinin antigen gene.

XX Surface protein; Mycobacterium bovis; BCG; adhesion; epithelium; ds;  
 KW Bacille Calmette-Guerin; Mycobacterium tuberculosis; epithelial cell;  
 KW heparin-binding haemagglutinin antigen; PCR; primer; amplification;  
 KW probe; hybridisation; chromosome; vaccine; diagnosis; immunoassay.

XX Mycobacterium bovis.

XX Key Location/Qualifiers

FT CDS 331..930

FT /\*tag= a

FT /product= "HBHA"

FT /note= "heparin-binding haemagglutinin antigen"

FT repeat\_region 811..864

FT /\*tag= b

FT /rpt\_unit= 811..828

FT repeat\_region 865..915

FT /\*tag= d

FT /rpt\_unit= 865..885

XX WO9744463-A2.

XX 27-NOV-1997.

XX 20-MAY-1997; 97WO-FR000896.

XX 17-MAY-1996; 96FR-00006168.

XX (INSP ) INST PASTEUR LILLE.

PA (INRM ) INST NAT SANTE & RECH MEDICALE.

XX Menozzi F, Loch C;

XX WPI; 1998-018517/02.

XX P-PSDB; AA744936.

XX New peptide(s) involved in adhesion of mycobacteria to epithelial cells -  
 PT used in vaccines and for diagnosing mycobacterial infection, also use of

PT sulphated saccharide(s) and glyco-conjugate(s) to prevent mycobacterial  
 PT adhesion.

XX Claim 21; Fig 10; 52pp; French.

XX This sequence represents the gene encoding a surface protein found on  
 CC Mycobacterium bovis BCG (Bacille Calmette-Guerin) or M. tuberculosis,  
 CC which enables mycobacteria to adhere to host, especially epithelial,  
 CC cells. The protein was isolated by passing a culture of BCG  
 CC microorganisms over a Heparin-Sepharose column and eluting proteins with  
 CC a gradient of 0-0.5 M NaCl. Fractions were separated by gel  
 CC electrophoresis and a 28 kD heparin binding protein purified. N-terminal  
 CC and internal peptide fractions were sequenced and the amino acid  
 CC sequences used to design PCR primers (AA76949-T76952). These amplified a  
 CC 150 bp fragment (AA76947) used as a probe to isolate chromosomal  
 CC fragments containing the gene which encodes a heparin-binding  
 CC haemagglutinin antigen (HBHA). The protein or peptides derived from it  
 CC (e.g. AA76949-T76952), can be used in vaccines against mycobacterial infections  
 CC (particularly M. bovis or M. tuberculosis), or can be used to diagnose  
 CC mycobacterial infection (by immunoassay detection of anti-HBHA  
 CC antibodies)

XX Sequence 1097 BP; 236 A; 324 C; 365 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 1097;

Best Local Similarity 100.0%; Pred. No. 0.035;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTCGAGCCGACTACTTCTGGGT 26

DB 944 GATTCGAGCCGACTACTTCTGGGT 919

## RESULT 3

ABX09143/c

ID ABX09143 standard; DNA; 86114 BP.

XX ABX09143;

AC ABX09143;

XX 08-APR-2003 (first entry)

XX Mycobacterium tuberculosis H37Rv BAC clone BAC-RV265.

XX Mycobacterioses; survival; virulence; protective antigen; vaccine;

XX Mycobacterial disease; tuberculosis; leprosy; ds; cosmid.

XX Mycobacterium tuberculosis.

OS WO200274903-A2.

XX 26-SEP-2002.

XX 22-FEB-2002; 2002WO-IB001973.

XX 22-FEB-2001; 2001US-0270123P.

XX (INSP ) INST PASTEUR.

XX Cole S;

XX WPI; 2002-759885/82.

XX Identifying and selecting genes for survival or virulence of mycobacteria  
 PT by a comparative genomic analysis of the sequences of Mycobacterium  
 PT tuberculosis and M. leprae.  
 XX Disclosure; Fig 7; 874pp; English.

XX This invention relates to a novel method for identifying essential genes  
 CC for survival or virulence of mycobacteria species. The method comprises  
 CC aligning the genomic sequence of a first mycobacteria species on a  
 CC genomic sequence of a second mycobacteria species and selecting a  
 CC polynucleotide sequence that is highly conserved in both genomes with no



CC counterparts in other bacterial genomic sequences and that corresponds to  
CC an essential gene for the survival or virulence of mycobacterium species.  
CC The method of the invention is useful for detecting M. tuberculosis or M.  
CC leprae infection. The method reduces the number of potential new targets  
CC and protective antigens for new drugs and vaccine compositions to treat  
CC and prevent mycobacterial diseases, particularly tuberculosis and  
CC leprosy. The present sequence represents a Mycobacterium cosmid DNA  
CC sequence used in the method of the invention

XX SQ Sequence 86114 BP; 14693 A; 27103 C; 28984 G; 15334 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 6; Length 86114;

Best Local Similarity 100.0%; Pred. No. 0.051;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTCGAGCCGCGACTACTTCTGGGT 26

Db 53007 GATTCGAGCCGCGACTACTTCTGGGT 52982

## RESULT 4

AAI99682\_05/c  
Continuation (6 of 45) of AAI99682 from base 500001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP Fragment Name Begin End

WP AAI99682\_01 1 110000

WP AAI99682\_02 100001 210000

WP AAI99682\_03 200001 310000

WP AAI99682\_04 300001 410000

WP AAI99682\_05 400001 510000

WP AAI99682\_06 500001 610000

WP AAI99682\_07 600001 710000

WP AAI99682\_08 700001 810000

WP AAI99682\_09 800001 910000

WP AAI99682\_10 900001 1010000

WP AAI99682\_11 1000001 1110000

WP AAI99682\_12 1100001 1210000

WP AAI99682\_13 1200001 1310000

WP AAI99682\_14 1300001 1410000

WP AAI99682\_15 1400001 1510000

WP AAI99682\_16 1500001 1610000

WP AAI99682\_17 1600001 1710000

WP AAI99682\_18 1700001 1810000

WP AAI99682\_19 1800001 1910000

WP AAI99682\_20 1900001 2010000

WP AAI99682\_21 2000001 2110000

WP AAI99682\_22 2100001 2210000

WP AAI99682\_23 2200001 2310000

WP AAI99682\_24 2300001 2410000

WP AAI99682\_25 2400001 2510000

WP AAI99682\_26 2500001 2610000

WP AAI99682\_27 2600001 2710000

WP AAI99682\_28 2700001 2810000

WP AAI99682\_29 2800001 2910000

WP AAI99682\_30 2900001 3010000

WP AAI99682\_31 3000001 3110000

WP AAI99682\_32 3100001 3210000

WP AAI99682\_33 3200001 3310000

WP AAI99682\_34 3300001 3410000

WP AAI99682\_35 3400001 3510000

WP AAI99682\_36 3500001 3610000

WP AAI99682\_37 3600001 3710000

WP AAI99682\_38 3700001 3810000

WP AAI99682\_39 3800001 3910000

WP AAI99682\_40 3900001 4010000

WP AAI99682\_41 4000001 4110000

WP AAI99682\_42 4100001 4210000

WP AAI99682\_43 4200001 4310000

WP AAI99682\_44 4300001 4410000

WP AAI99682\_45 4400001 441529

Query Match

Best Local Similarity

100.0%;

Pred. No. 0.052;

Score 26; DB 4; Length 110000;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTCGAGCCGCGACTACTTCTGGGT 26

Db 66408 GATTCGAGCCGCGACTACTTCTGGGT 66383

## RESULT 5

AAI99683\_05/c  
Continuation (6 of 44) of AAI99683 from base 500001 (Mycobacterium tuberculosis strain I  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP Fragment Name Begin End

WP AAI99683\_01 1 110000

WP AAI99683\_02 100001 210000

WP AAI99683\_03 200001 310000

WP AAI99683\_04 300001 410000

WP AAI99683\_05 400001 510000

WP AAI99683\_06 500001 610000

WP AAI99683\_07 600001 710000

WP AAI99683\_08 700001 810000

WP AAI99683\_09 800001 910000

WP AAI99683\_10 900001 1010000

WP AAI99683\_11 1000001 1110000

WP AAI99683\_12 1100001 1210000

WP AAI99683\_13 1200001 1310000

WP AAI99683\_14 1300001 1410000

WP AAI99683\_15 1400001 1510000

WP AAI99683\_16 1500001 1610000

WP AAI99683\_17 1600001 1710000

WP AAI99683\_18 1700001 1810000

WP AAI99683\_19 1800001 1910000

WP AAI99683\_20 1900001 2010000

WP AAI99683\_21 2000001 2110000

WP AAI99683\_22 2100001 2210000

WP AAI99683\_23 2200001 2310000

WP AAI99683\_24 2300001 2410000

WP AAI99683\_25 2400001 2510000

WP AAI99683\_26 2500001 2610000

WP AAI99683\_27 2600001 2710000

WP AAI99683\_28 2700001 2810000

WP AAI99683\_29 2800001 2910000

WP AAI99683\_30 2900001 3010000

WP AAI99683\_31 3000001 3110000

WP AAI99683\_32 3100001 3210000

WP AAI99683\_33 3200001 3310000

WP AAI99683\_34 3300001 3410000

WP AAI99683\_35 3400001 3510000

WP AAI99683\_36 3500001 3610000

WP AAI99683\_37 3600001 3710000

WP AAI99683\_38 3700001 3810000

WP AAI99683\_39 3800001 3910000

WP AAI99683\_40 3900001 4010000

WP AAI99683\_41 4000001 4110000

WP AAI99683\_42 4100001 4210000

WP AAI99683\_43 4200001 4310000

WP AAI99683\_44 4300001 4403765

Query Match 100.0%; Score 26; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0.052;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTCGAGCCGCGACTACTTCTGGGT 26

Db 67850 GATTCGAGCCGCGACTACTTCTGGGT 67825

## RESULT 6

AAA82062/c

ID AAA82062 standard; DNA; 493 BP.

XX AC AAA82062;

XX XX

DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_609 SEQ ID NO:609.  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 XX antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX Neisseria meningitidis.  
 OS Neisseria meningitidis.  
 XX WO200022430-A2.  
 XX 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US023573.  
 XX 09-OCT-1998; 98US-0103794P.  
 XX 30-APR-1999; 99US-0132068P.  
 XX (CHIR) CHIRON CORP.  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC, Scarlato V;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappelli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be used  
 PT in the diagnosis and treatment of N. meningitidis infection and other  
 PT Neisseria infections, for example, N. gonorrhoea.  
 XX Claim 7; Page 1669; 1760pp; English.  
 XX The present invention describes methods of obtaining immunogenic proteins  
 CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent  
 CC specifically claimed Neisseria meningitidis genomic DNA sequences;  
 CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA  
 CC sequences and their corresponding proteins; AAA81254 to AAA81259 and  
 CC AAA81304 to AAA81321 represent PCR primers used in the isolation of  
 CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent  
 CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all  
 CC used in the exemplification of the present invention. The nucleic acid  
 CC sequences, protein sequences, and antibodies against them, can be used in  
 CC the manufacture of a composition. The composition can be used as a  
 CC medicament (or in the manufacture of a medicament) for treating,  
 CC preventing or diagnosing infection due to Neisseria bacteria. For  
 CC example, some of the identified proteins could be components of vaccines  
 CC against Meningococcus B; against all serotypes; and/or against all  
 CC pathogenic Neisseriae. Identification of sequences from the bacterium  
 CC will also facilitate production of biological probes, particularly  
 CC organism-specific probes. Attempts to make efficacious Meningococcus B  
 CC vaccines have failed mainly due to antigen tolerance. Multivalent  
 CC vaccines have also been tried but none have successfully overcome  
 CC antigenic variability. The provision of further, complete sequences may  
 CC provide an opportunity to identify secreted or surface exposed proteins  
 CC that may be presumed targets for the immune system and which are not  
 CC antigenically variable or at least more conserved than other more  
 CC variable regions  
 XX Sequence 493 BP; 92 A; 127 C; 116 G; 158 T; 0 U; 0 Other;  
 SQ Query Match 70.0%; Score 18.2; DB 3; Length 493;  
 Best Local Similarity 87.0%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GATTCGACCCGACTACTTCG 23  
 DB 279 GATCTGGAGCCGACTACTTCG 257  
 RESULT 7  
 ADC87234  
 ID ADC87234 standard; DNA; 21581 BP.  
 XX

AC ADC87234;  
 XX 01-JAN-2004 (first entry)  
 XX Human GPCR gene SEQ ID NO:1687.  
 DE ds; gene; human; GPCR;  
 KW guanosine triphosphate-binding protein coupled receptor; gene therapy.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX EP1270724-A2.  
 XX 02-JAN-2003.  
 XX 18-JUN-2002; 2002EP-00013517.  
 XX 18-JUN-2001; 2001JP-00246789.  
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;  
 XX WPI; 2003-315783/31.  
 XX P-PSDB; ADC87235.  
 XX New polynucleotide, useful for preparing a composition for treating a  
 PT patient in need of increased or suppressed activity or expression of the  
 PT guanosine triphosphate-binding protein coupled receptor.  
 XX Claim 1; SEQ ID NO 1687; 28pp; English.  
 XX The invention relates to a novel polynucleotide encoding a guanosine  
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
 CC the invention may have a use in gene therapy. The polynucleotide and  
 CC polypeptide are useful for preparing a composition for treating a patient  
 CC in need of increased or suppressed activity or expression of the  
 CC guanosine triphosphate-binding protein coupled receptor. The  
 CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
 XX invention.  
 XX Sequence 21581 BP; 5131 A; 5003 C; 5838 G; 5609 T; 0 U; 0 Other;  
 SQ Query Match 69.2%; Score 18; DB 9; Length 21581;  
 Best Local Similarity 80.8%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 GATTCGGAGCCCGACTACTTCTGGGT 26  
 DB 647 GATTCGGAGTCTGGCTGCTCCTGGGT 672  
 RESULT 8  
 ADB10175/c  
 ID ADB10175 standard; DNA; 288 BP.  
 XX ADB10175;  
 AC ADB10175;  
 XX 20-NOV-2003 (first entry)  
 XX Alloiococcus otitis antigenic protein encoding DNA SEQ ID NO:4887.  
 DE Alloiococcus otitis antigenic protein; immunogenic; immunisation;  
 KW Alloiococcus otitis; antigenic protein; immunogenic; infection; gene  
 KW gene therapy; gram-positive bacterium; infection; gene; ds.  
 XX Alloiococcus otitis.  
 OS Alloiococcus otitis.  
 XX WO2003048304-A2.  
 XX 12-JUN-2003.  
 XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.  
PR 18-NOV-2002; 2002US-0426742P.  
XX (AMHP ) WYETH HOLDINGS CORP.  
XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;  
XX WPI; 2003-505284/47.  
XX DR P-PSDB; ADB10172.  
XX  
XX New Alloiococcus otitidis polynucleotides and polypeptides, useful for  
PT treating and diagnosing diseases, drug screening assays and monitoring of  
PT effects during drug clinical trials.  
XX  
XX Claim 7; SEQ ID NO 4887; 1019pp; English.  
XX  
XX The present invention describes an isolated polynucleotide (I) of  
CC Alloiococcus otitidis genomic DNA, which encodes an antigenic protein.  
CC Alloiococcus otitidis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
CC expression vector comprising the novel isolated polynucleotide (I), its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunizing against Alloiococcus otitidis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying Alloiococcus  
CC otitidis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (I) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Alloiococcus  
CC otitidis. The present sequence encodes an Alloiococcus otitidis antigen  
CC protein from the present invention.  
XX  
XX Sequence 288 BP; 99 A; 49 C; 54 G; 76 T; 0 U; 0 Other;  
SQ  
Query Match 67.7%; Score 17.6; DB 8; Length 288;  
Best Local Similarity 83.3%; Pred. NO. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ATTCGAGCCGACACTTCTGGG 25  
DB 224 ATTGGTCCCGCTCTCTCTGGG 201  
RESULT 9  
ABL08045.  
ID ABL08045 standard; cDNA; 4268 BP.  
XX ABL08045;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 18617.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200117042-A2.  
XX  
XX 27-SEP-2001.  
PD

XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB63942.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 18617; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
XX Sequence 4268 BP; 1117 A; 1157 C; 1128 G; 866 T; 0 U; 0 Other;  
SQ  
Query Match 67.7%; Score 17.6; DB 4; Length 4268;  
Best Local Similarity 83.3%; Pred. NO. 3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GATTCGAGCCGACACTTCTGGG 24  
DB 3990 GTTTCGGTGACGACGCTCTCTGGG 4013  
RESULT 10  
AAL37093  
ID AAL37093 standard; DNA; 5227 BP.  
XX  
XX AAL37093;  
XX  
XX 08-JAN-2002 (first entry)  
XX  
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3458.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155367-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001338.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 15-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR

us-10-624-714-8.rng

Mon May 3 09:03:27 2004

PR	19-MAY-2000;	2000US-0205515P.	PR	20-OCT-2000;	2000US-0241221P.
PR	07-JUN-2000;	2000US-0209467P.	PR	20-OCT-2000;	2000US-0241785P.
PR	28-JUN-2000;	2000US-0214886P.	PR	20-OCT-2000;	2000US-0241786P.
PR	30-JUN-2000;	2000US-0215135P.	PR	20-OCT-2000;	2000US-0241787P.
PR	07-JUL-2000;	2000US-0216647P.	PR	20-OCT-2000;	2000US-0241808P.
PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241809P.
PR	11-JUL-2000;	2000US-0217487P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUL-2000;	2000US-0217496P.	PR	01-NOV-2000;	2000US-0244617P.
PR	14-JUL-2000;	2000US-0218290P.	PR	08-NOV-2000;	2000US-0246474P.
PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226686P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229503P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250160P.
PR	12-SEP-2000;	2000US-0231966P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232319P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0233393P.	PR	06-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0233400P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0233401P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234224P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234597P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	FA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327P.	XX	WPI; 2001-451937/48.	
PR	29-SEP-2000;	2000US-0236367P.	DR		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	PT	Isolated polypeptide for treating, preventing and/ or prognosing	
PR	29-SEP-2000;	2000US-0236370P.	PT	disorders related to the musculoskeletal system including musculoskeletal	
PR	02-OCT-2000;	2000US-0236802P.	XX	cancers and also for testing and detection e.g. diagnosis.	
PR	02-OCT-2000;	2000US-0237037P.	PS	Example 2; SEQ ID NO 3458; 781pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	CC	The invention relates to novel genes (AAL34669-AAL37666) and proteins	
PR	02-OCT-2000;	2000US-0237040P.	CC	(ABB03087-ABB04109) associated with the musculoskeletal system useful for	
PR	13-OCT-2000;	2000US-0239935P.	CC	preventing, treating or ameliorating medical conditions e.g. by protein	
PR	13-OCT-2000;	2000US-0239937P.	CC	or gene therapy. The genes are isolated from a range of human tissues	
PR	20-OCT-2000;	2000US-0240960P.	CC		

CC disclosed in the specification. The nucleic acids, proteins, antibodies, and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 5227 BP; 1522 A; 997 C; 988 G; 1720 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 4; Length 5227;

Best Local Similarity 83.3%; Pred. No. 3e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ATTCGGACCCGACTACTCTCTGGG 25

DB 1819 AGTCGGAGCCTCACTCTCTGGG 1842

RESULT 11

ABAL4606

ID ABAL4606 standard; DNA; 5227 BP.

XX AC ABAL4606;

XX DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polynucleotide SEQ ID NO 6937.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide; immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary; antiparkinsonian; antiskickling; antianaemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antilulcer; anticonvulsant; antifungal; antiparasitic; cardiant; immune disorder; cardiovascular disorder; neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX OS Homo sapiens.

XX PN WO200159063-A2.

XX PD 16-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001334.

XX PR 31-JAN-2000; 2000US-01790652P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226968P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228242P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 23-SEP-2000; 2000US-0234977P.

PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0242221P.

PR 01-NOV-2000; 2000US-024617P.

PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.

PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating nervous system cancers  
XX and metastases.  
XX  
XX Disclosure; SEQ ID NO 6937; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
XX (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX infectious diseases such as viral, bacterial, fungal and parasitic  
XX infections. Note: The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SEQ Sequence 5227 BP; 1522 A; 997 C; 988 G; 1720 T; 0 U; 0 Other;  
Query Match 67.7%; Score 17.6; DB 5; Length 5227;  
Best Local Similarity 83.3%; Pred. No. 3e-02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 ATTGGAGCCGACTACTCTCTGGG 25  
Db 1819 AGTCGGAGCCTCACTCTCTGGG 1842  
RESULT 12  
ID ABA14501 standard; DNA; 5227 BP.  
XX AC ABA14501;  
XX 23-JAN-2002 (first entry)  
XX Human nervous system related polynucleotide SEQ ID NO 6832.  
XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;  
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
XX antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX Homo sapiens.  
XX WO200159063-A2.  
XX 16-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001334.  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184564P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 11-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 14-JUL-2000; 2000US-0217496P.  
XX 26-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 23-AUG-2000; 2000US-0227182P.  
XX 30-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.



AC ABAL7614;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 9945.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;  
KW antineumatic; hepatocytic; cerebroprotective; antiinflammatory;  
KW antiallergic; antiblastic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
PN  
XX  
XX 16-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001334.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198121P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217498P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225211P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229511P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 13-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 28-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0242211P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.



PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251980P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX  
XX Disclosure; SEQ ID NO 9945; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5227 BP; 1522 A; 997 C; 985 G; 1720 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 5; Length 5227;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ATTCGGAGCCGACTACTTCGGG 25  
Db 1819 AGTCGGAGCCCTCCTTCGGG 1842

RESULT 14  
ABX60081  
ID ABX60081 standard; cDNA; 5227 BP.  
XX  
XX AC ABX60081;  
XX  
DT 26-FEB-2003 (first entry)  
XX  
DE cDNA encoding novel human musculoskeletal system antigen #2425.  
XX  
XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
KW cardiovascular condition; wound; injury; burn; angio genesis; ulcer;  
KW post-operative tissue repair; limb regeneration; neuronal growth;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW AIDS-related complex; chondrocyte growth; bone regeneration;  
KW Periodontal regeneration; tissue transport; bone graft; skin aging;  
KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
KW cell growth; organ transplant; cell differentiation; body height; weight;

KW hair colour; eye colour; skin; percentage of adipose tissue;  
KW pigmentation; cosmetic surgery; metabolism; biorhythm; circadian rhythm;  
KW depression; tendency for violence; pain; reproductive capability;  
KW hormone level; endocrine level; appetite; libido; memory; stress;  
KW storage capability; fat content; lipid content; protein content;  
KW carbohydrate content; vitamin content; cofactor content;  
KW nutritional component.  
XX  
XX Homo sapiens.  
XX  
XX US2002147140-A1.  
XX  
XX 10-OCT-2002.  
XX  
XX 17-JAN-2001; 2001US-00764877.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220363P.  
PR 28-JUL-2000; 2000US-0220364P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 03-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236302P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
XX (ROSE/) ROSEN C A.  
XX (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
XX

DR WPI; 2003-128199/12.  
XX  
PT Isolated nucleic acid molecules encoding musculoskeletal system  
PT associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
PS Disclosure; SEQ ID NO 3458; 321pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule comprising a  
CC sequence encoding musculoskeletal system associated polypeptides useful  
CC for detecting disorders, e.g., cancer or cancer metastases, in animals or  
CC humans. The nucleic acid; stimulates re-vascularisation of ischemic  
CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
CC and other cardiovascular conditions; treats wounds due to injuries,  
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
CC and limb regeneration; stimulates neuronal growth; can treat and prevent  
CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or  
CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, since RGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, haematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, eye colour,  
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, circadian rhythms,  
CC depression tendency for violence, tolerance for pain, reproductive  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?docID=20020147140  
XX  
SQ Sequence 5227 BP; 1522 A; 997 C; 988 G; 1720 T; 0 U; 0 Other;  
Query Match 67.7%; Score 17.6; DB 7; Length 5227;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ATTCGGAGCCCGACTACTTCTGGG 25  
DB 1819 AGTCGGAGCCCTCCTCTCTGGG 1842  
RESULT 15  
ABL08044/c  
ID ABL08044 standard; cDNA; 6670 BP.  
XX  
AC ABL08044;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18614.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
OS Drosophila melanogaster.  
XX  
PN WC200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers BW;  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB63941.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 18614; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 6670 BP; 1632 A; 1623 C; 1679 G; 1736 T; 0 U; 0 Other;  
Query Match 67.7%; Score 17.6; DB 4; Length 6670;  
Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GATTCGGAGCCCGACTACTTCTGGG 24  
DB 1337 GTTTCGGTGCACGACTGCTTCTGG 1314  
Search completed: April 29, 2004, 04:57:09  
Job time : 217.727 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 2050.13 Seconds  
(without alignments)  
378.717 Million cell updates/sec

Title: US-10-624-714-8

Perfect score: 26  
Sequence: 1 gattccggagccgactactcttgggt 26

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hrc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hrc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_ham:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl:\*

29: gb\_gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.6	75.4	244	10 BF853535	BF853535 MR2-EN009
C 2	19.2	73.8	544	14 CA722298	CA722298 wdsic.pk0
C 3	19.2	73.8	1103	12 BG333879	BG333879 602460211
C 4	18.8	72.3	405	12 BG988570	BG988570 CM3-HT002

5	18.8	72.3	586	13	BU480799
6	18.8	72.3	587	9	AL119975
C 7	18.8	72.3	628	10	BE622229
C 8	18.8	72.3	712	10	BE896229
C 9	18.8	72.3	742	12	BG763564
C 10	18.8	72.3	833	12	BI090351
C 11	18.8	72.3	875	12	BI830850
C 12	18.8	72.3	983	13	EX460914
C 13	18.8	72.3	1003	13	EX334881
C 14	18.8	72.3	1070	12	BM562054
C 15	18.8	72.3	1073	11	BC016935
C 16	18.8	72.3	1119	12	BM471401
C 17	18.8	72.3	1201	9	AL532474
C 18	18.8	72.3	1201	13	EX382423
C 19	18.8	72.3	1201	13	EX427277
C 20	18.8	72.3	3127	11	BC015800
C 21	18.6	71.5	498	12	BI982766
C 22	18.6	71.5	522	12	BG279396
C 23	18.6	71.5	668	29	AG086843
C 24	18.2	70.0	359	14	CA929359
C 25	18.2	70.0	387	14	CA929302
C 26	18.2	70.0	516	9	AV437656
C 27	18.2	70.0	676	10	BF864410
C 28	18.2	70.0	774	28	AQ632881
C 29	18.2	70.0	774	28	AQ632881
C 30	18.2	70.0	774	28	AQ632881
C 31	18.2	70.0	774	28	AQ632881
C 32	18.2	70.0	774	28	AQ632881
C 33	18.2	70.0	774	28	AQ632881
C 34	18.2	70.0	774	28	AQ632881
C 35	18.2	70.0	774	28	AQ632881
C 36	18.2	70.0	774	28	AQ632881
C 37	18.2	70.0	774	28	AQ632881
C 38	18.2	70.0	774	28	AQ632881
C 39	18.2	70.0	774	28	AQ632881
C 40	18.2	70.0	774	28	AQ632881
C 41	18.2	70.0	774	28	AQ632881
C 42	18.2	70.0	774	28	AQ632881
C 43	18.2	70.0	774	28	AQ632881
C 44	18.2	70.0	774	28	AQ632881
C 45	18.2	70.0	774	28	AQ632881

#### ALIGNMENTS

RESULT 1  
BF853535/c  
LOCUS MR2-EN0093-211200-002-g05 EN0093 Homo sapiens cdNA, mRNA sequence.  
DEFINITION MR2-EN0093-211200-002-g05 EN0093 Homo sapiens cdNA, mRNA sequence.  
ACCESSION BF853535  
VERSION BF853535.1 GI:12241279  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 244)  
AUTHORS Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F., Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R., Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H., Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and Simpson A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

BU480799 603472063  
AL119975 DKFZP761E  
BE622229 601440794  
BE896229 601512650  
BG763564 602735892  
BI090351 602855465  
BI830850 603073190  
EX460914 EX460914  
EX334881 EX334881  
BM562054 AGENCOURT  
BC016935 Homo sapi  
BM471401 AGENCOURT  
AL532474 AL532474  
BX382423 BX382423  
EX427277 EX427277  
BC015800 Homo sapi  
BI982766 fu57h12.y  
BG279396 b2h03np.r  
AG086843 Pan trogl  
CA929359 MTU2CA.PI  
CA929302 MTU2CA.PI  
AV437656 AV437656  
BF864410 963051E04  
AQ632881 RPCI-11-4  
BM722068 UI-E-EOO-  
AQ686841 nbe0030D  
BQ758827 Ema07 SQ  
BU950352 i080a10.x  
BM352717 ig70509.x  
BX245916 Dario rer  
BM353680 i958f08.x  
CC833378 ZMMBB018  
BM373454 Ema04 SO  
BE37762 601217560  
BZ198017 CH230-416  
CC679641 OGUKA34TV  
BE422199 HWM021CD  
AA040496 zf05g08.i  
AL900544 AL900544  
CF217418 AGENCOURT

## Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR2&t2=MR2-EN0093-211200-002-g05&t3=2000-12-21&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 200.

## FEATURES

source

1..244  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="EN0093"

/note="Organ: lung normal; Vector: puc18; Site: SmaI; Site: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 75.4%; Score 19.6; DB 10; Length 244;  
Best Local Similarity 84.6%; Pred. No. 4.9e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCCGGACTACTTCTGGGT 26

Db 66 GAGTGGAGCCCGCCACACTTCTGGGT 41

## RESULT 2

CA722298

LOCUS

CA722298 wdsic.pk003.j4.f wdsic Triticum aestivum cDNA clone 544 bp mRNA linear EST 26-NOV-2002

CA722298 wdsic.pk003.j4.f 3' end, mRNA sequence.

CA722298 EST.

CA722298 1 GI:254444091

Triticum aestivum (bread wheat)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

REFERENCE

AUTHORS

Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,

Miao, G., Caraher, N., and Hanafey, M.K.

DuPont Wheat cDNA Sequence

Unpublished (2002)

Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: T7.

## FEATURES

source

1..544  
Location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="monococcum"  
/db\_xref="taxon:4565"  
/clone="wdsic.pk003.j4.f"  
/tissue\_type="seedling"  
/lab\_host="DH10B"  
/note="Vector: pBluescript SK+; Site: EcoRI; Site: 2; XhoI; Wheat (Triticum aestivum, monococcum) developing

## ORIGIN

seedling"

Query Match 73.8%; Score 19.2; DB 14; Length 544;  
Best Local Similarity 87.5%; Pred. No. 9.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TTCGGAGCCCGGACTACTTCTGGGT 26

Db 378 TTGGGACCCGACCACTTCTGGGT 401

## RESULT 3

BG333879

LOCUS

BG333879 602460211F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4576899 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1CM1290 row: f column: 04

High quality sequence stop: 1.

Location/Qualifiers

1..1103

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4576899"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 20"

/note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site: 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected 500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 73.8%; Score 19.2; DB 12; Length 1103;  
Best Local Similarity 87.5%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TTCGGAGCCCGGACTACTTCTGGGT 26

Db 926 TTCGGGCGCGGCTACTTCTGGGT 949

## RESULT 4

BG988570/c

LOCUS

BG988570 CM3-HT0023-180101-638-h05 HT0023 Homo sapiens cDNA, mRNA sequence. EST 13-JUN-2001

ACCESSION

VERSION

KEYWORDS EST. Homo sapiens (human)

SOURCE MEDLINE PUBMED COMMENT

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 405)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brownstein,A., Geolvilleira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-HT0023-180101-638-h05&t3=2001-01-18&t4=1>) Seq primer: puc 18 forward High quality sequence stop: 404. Location/Qualifiers 1..405 /organism="Homo sapiens" /mol\_type="mRNA" /db\_xref="taxon:9606" /dev\_stage="adult" /clone\_lib="HRC023" /note="Organ: head neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES source

1..586 /organism="Gallus gallus" /mol\_type="mRNA" /strain="Layer and broiler" /db\_xref="taxon:9031" /clone="ChEST349n15" /sex="Male and female" /tissue\_type="Chondrocytes isolated from growth plate cartilage" /dev\_stage="adult" /lab\_host="DH10B" /clone\_lib="CSEQRN22" /note="Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI. This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., FNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 72.3%; Score 18.8; DB 13; Length 586; Best Local Similarity 90.9%; Pred. No. 1.4e+03; Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATTCGAGCCGCGACTACTTCTG 23

DB 92 ACTCGTGCCGCGACTACTTCTG 113

RESULT 6

AL119975/c

LOCUS DXFZP761E222\_r1 761 (synonym: hamy2) Homo sapiens cDNA clone

DEFINITION DXFZP761E222 5', mRNA sequence.

ACCESSION AL119975

VERSION AL119975

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 587)

AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and Wiemann,S.

TITLE EST (Ottenwaelder, et al.)

JOURNAL Unpublished (1999)

COMMENT Contact: MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinsried/Germany) within the cDNA

KEYWORDS EST. Homo sapiens (human)

SOURCE MEDLINE PUBMED COMMENT

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 405)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brownstein,A., Geolvilleira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-HT0023-180101-638-h05&t3=2001-01-18&t4=1>) Seq primer: puc 18 forward High quality sequence stop: 404. Location/Qualifiers 1..405 /organism="Homo sapiens" /mol\_type="mRNA" /db\_xref="taxon:9606" /dev\_stage="adult" /clone\_lib="HRC023" /note="Organ: head neck; Vector: puc18; Site: 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 72.3%; Score 18.8; DB 12; Length 405; Best Local Similarity 90.9%; Pred. No. 1.3e+03; Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGAGCCGCGACTACTTCTGGG 25

DB 78 TCGAGCCGCGACTCTCTCAGGG 57

RESULT 5

BU480799

LOCUS 603472063F1 CSEQRN22 Gallus gallus cDNA clone ChEST349n15 5', mRNA

DEFINITION sequence.

ACCESSION BU480799

VERSION BU480799

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 586)

AUTHORS Boardman,P.B., Sanz-Esquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

TITLE A Comprehensive Collection of Chicken cDNAs

sequencing consortium of the German Genome Project. No's 1 sequence available.  
This clone (DKFZp761E222) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source  
1..587  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp761E222"  
/tissue\_type="amygdala"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_72"  
/note="Vector: pSPori; Site\_1: NotI; Site\_2: SalI"

## ORIGIN

Query Match 72.3%; Score 18.8; DB 9; Length 587;  
Best Local Similarity 90.9%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY

4 TCGAGCCCGGACTACTCTGGG 25

## Db

213 TCGAGCCCGGACTCTTCAGG 192

## RESULT 7

BE622229/c  
LOCUS BE622229 628 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601440794F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3915791 5', mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 628)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTP/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9735 row: k column: 24  
High quality sequence stop: 619.

## FEATURES

source  
1..628  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3915791"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_72"  
/note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

## ORIGIN

Query Match 72.3%; Score 18.8; DB 10; Length 628;  
Best Local Similarity 90.9%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY

4 TCGAGCCCGGACTACTCTGGG 25

## Db

587 TCGAGCCCGGACTCTTCAGG 566

## RESULT 8

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

Tissue Procurement: ATCC/DCTD/DTP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM1719 row: d column: 22  
 High quality sequence stop: 618.  
 Location/Qualifiers  
 1. .742  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4861029"  
 /tissue\_type="melanotic melanoma, high MDR (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_49"  
 /note="Organ: Skin; Vector: pOTB7; Site:1: XhoI; Site:2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library. |"

## ORIGIN

Query Match 72.3%; Score 18.8; DB 12; Length 742;  
 Best Local Similarity 90.9%; Pred. No. 1.5e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 TCGGAGCCCGACTACTTCTGGG 25  
 |||||  
 Db 209 TCGGAGCCCGACTCTCTCAGGG 188

## RESULT 10

BI090351/c  
 LOCUS 833 bp mRNA linear EST 20-JUN-2001  
 DEFINITION 602855465F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4996528 5',  
 mRNA sequence.  
 ACCESSION BI090351  
 VERSION BI090351.1 GI:14508681  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 833)  
 NIH-MGC http://img.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ARCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

## FEATURES

source  
 1. .833  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4996528"  
 /cell\_line="MGC36"  
 /lab\_host="DH10B"

/clone\_lib="NIH\_MGC\_10"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.5 Kb. Library prepared by Life  
 Technologies."

## ORIGIN

Query Match 72.3%; Score 18.8; DB 12; Length 833;  
 Best Local Similarity 90.9%; Pred. No. 1.6e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 TCGGAGCCCGACTACTTCTGGG 25  
 |||||  
 Db 174 TCGGAGCCCGACTCTCTCAGGG 153

## RESULT 11

BI830650/c  
 LOCUS 875 bp mRNA linear EST 04-OCT-2001  
 DEFINITION 603073190F1 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5165134 5',  
 mRNA sequence.  
 ACCESSION BI830650  
 VERSION BI830650.1 GI:15942200  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 875)  
 NIH-MGC http://img.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11410 row: c column: 23  
 High quality sequence stop: 873.  
 Location/Qualifiers  
 1. .875  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5165134"  
 /tissue\_type="medulla"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_119"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 Kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 013. Note:  
 this is a NIH\_MGC Library."

## FEATURES

source  
 1. .875  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5165134"  
 /tissue\_type="medulla"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_119"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 Kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 013. Note:  
 this is a NIH\_MGC Library."

## ORIGIN

Query Match 72.3%; Score 18.8; DB 12; Length 875;  
 Best Local Similarity 90.9%; Pred. No. 1.6e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 TCGGAGCCCGACTACTTCTGGG 25  
 |||||  
 Db 674 TCGGAGCCCGACTCTCTCAGGG 653

## RESULT 12

```

BX460914/c    BX460914    983 bp    mRNA    linear    EST 22-MAY-2003
LOCUS          BX460914    Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION     CS0DF020YD06 5-PRIME, mRNA sequence.
ACCESSION      BX460914
VERSION        BX460914.1 GI:31021281
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 983)
AUTHORS        Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               Library was constructed by Life Technologies, a division of
               Invitrogen. This sequence belongs to sequence cluster 10726.f For
               more information about this cluster, see
               http://www.genoscope.cns.fr/
               cgi-bin/cluster.cgi?seq=CS0DF020YD06&cluster=10726.f. Contact :
               Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
               Faraday Avenue Genoscope sequence ID : CS0DF020YD03QP1.

FEATURES             Location/Qualifiers
     source           1..983
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF020YD06"
                     /tissue_type="FETAL BRAIN"
                     /dev stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."

ORIGIN
Query Match      72.3%; Score 18.8; DB 13; Length 983;
Best Local Similarity 90.9%; Pred.No. 1.6e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 TCGAGCCCGACTACTTCTGGG 25
        |||||
        722 TCGAGCCCGACTCTCTTCAGGG 701

Db

RESULT 13
BX334881/c    BX334881    1003 bp    mRNA    linear    EST 02-MAY-2003
LOCUS          BX334881    Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION     clone CS0DI008YG05 5-PRIME, mRNA sequence.
ACCESSION      BX334881
VERSION        BX334881.1 GI:30333361
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1003)
AUTHORS        Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               Library was constructed by Life Technologies, a division of

```

```

Invitrogen. This sequence belongs to sequence cluster 10726.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI008AD03QP1&cluster=10726.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI008AD03QP1.

FEATURES             Location/Qualifiers
     source           1..1003
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI008YG05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /notes="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoRV
                     sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      72.3%; Score 18.8; DB 13; Length 1003;
Best Local Similarity 90.9%; Pred.No. 1.7e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 TCGAGCCCGACTACTTCTGGG 25
        |||||
        738 TCGAGCCCGACTCTTCAGGG 717

Db

RESULT 14
BM562054/c    BM562054    1070 bp    mRNA    linear    EST 20-FEB-2002
LOCUS          BM562054    NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745415
DEFINITION     AGENCOURT 6562031 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745415
               5', mRNA sequence.
ACCESSION      BM562054
VERSION        BM562054.1 GI:18807813
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1070)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs@mail.nih.gov
               Tissue Procurement: Life Technologies, Inc.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM12768 row: f column: 08
               High quality sequence stop: 678.

FEATURES             Location/Qualifiers
     source           1..1070
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5745415"
                     /tissue_type="leukocyte"
                     /lab_host="DH108"
                     /clone_lib="NIH_MGC_118"
                     /notes="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV
                     (destroyed); RNA source leukocytes from anonymous pool of
                     non-activated adult donors. Library is oligo-dT primed
                     and directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.7 kb, insert size range
                     1.2-3.3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber

```



Job time : 2054.13 secs

(Invitrogen). Research Genetics tracking code 027. Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 72.3%; Score 18.8; DB 12; Length 1070;  
Best Local Similarity 90.9%; Pred. No. 1.7e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGAGCCCGACTCTCTGGG 25  
|||||  
DB 452 TCGAGCCCGACTCTCTCAGG 431

RESULT 15

BC016935/c  
LOCUS  
DEFINITION Homo sapiens, similar to DNA segment on chromosome X and Y (unique)  
155 expressed sequence, clone IMAGE:4430810, mRNA.

ACCESSION BC016935

VERSION BC016935.1 GI:16877369

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1073)

Strausberg, R.

Direct Submission

Submitted (05-NOV-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapps@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: http://www.shgc.stanford.edu

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov

Series: IRAK Plate: 27 Row: n Column: 4

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 10835221

This clone has the following problem: frame shifted.

FEATURES

source  
1..1073  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4430810"  
/tissue\_type="Prostate, adenocarcinoma."  
/clone\_lib="NIH MSC\_91"  
/lab\_host="DH10E"  
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 72.3%; Score 18.8; DB 11; Length 1073;  
Best Local Similarity 90.9%; Pred. No. 1.7e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCGAGCCCGACTCTCTGGG 25  
|||||  
DB 696 TCGAGCCCGACTCTCTCAGG 675

Search completed: April 29, 2004, 11:37:09

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 45.8824 Seconds  
(without alignments)  
314.472 Million cell updates/sec

Title: US-10-624-714-8

Perfect score: 26

Sequence: 1 gattcgagccgactactcttgggt 26

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	26	100.0	4403765	3	US-09-103-840A-2
C 2	26	100.0	441529	3	US-09-103-840A-1
C 3	17	65.4	209	3	US-09-063-893A-2
C 4	17	65.4	369	4	US-09-560-198A-3
C 5	17	65.4	3111	3	US-09-487-444-3
C 6	16.8	64.6	726	3	US-08-818-112-24
C 7	16.8	64.6	726	4	US-08-818-111-24
C 8	16.8	64.6	726	4	US-09-056-556-24
C 9	16.8	64.6	726	4	US-09-072-596-24
C 10	16.8	64.6	726	4	US-09-072-967-24
C 11	16.4	63.1	529	4	US-09-669-751-37
C 12	16.4	63.1	28804	2	US-08-592-874-1
C 13	16.4	63.1	28804	2	US-09-096-942-2
C 14	16.4	63.1	28804	3	US-09-096-867-2
C 15	16	61.5	516	4	US-09-252-991A-10247
C 16	16	61.5	528	4	US-09-833-381-1789
C 17	16	61.5	717	4	US-09-252-991A-10606
C 18	16	61.5	1094	3	US-09-441-346A-1
C 19	16	61.5	1094	4	US-09-789-453A-1
C 20	16	61.5	1659	4	US-09-252-991A-10461
C 21	16	61.5	1990	4	US-09-149-476-281
C 22	16	61.5	2041	4	US-09-149-476-131
C 23	16	61.5	2072	4	US-09-833-381-1781
C 24	16	61.5	3267	4	US-09-252-991A-10707
C 25	16	61.5	7577	4	US-09-637-048C-3
C 26	16	61.5	7621	4	US-09-637-048C-6
C 27	15.8	60.8	336	4	US-09-540-236-750

```

C 28 15.8 60.8 269223 4 US-09-596-002-41 Sequence 41, Appli
C 29 15.6 60.0 33 4 US-09-284-627-2 Sequence 2, Appli
C 30 15.6 60.0 465 4 US-09-252-991A-9706 Sequence 9706, Ap
C 31 15.6 60.0 747 4 US-09-252-991A-4328 Sequence 4328, Ap
C 32 15.6 60.0 849 4 US-09-252-991A-4526 Sequence 4526, Ap
C 33 15.6 60.0 1002 4 US-09-252-991A-4423 Sequence 4423, Ap
C 34 15.6 60.0 1087 4 US-09-372-422A-29 Sequence 29, Appli
C 35 15.6 60.0 1527 4 US-09-252-991A-9583 Sequence 9583, Ap
C 36 15.6 60.0 1677 4 US-09-252-991A-9544 Sequence 9544, Ap
C 37 15.6 60.0 2286 4 US-09-065-027-5 Sequence 5, Appli
C 38 15.6 60.0 2481 2 US-08-630-118A-1 Sequence 1, Appli
C 39 15.6 60.0 2481 2 US-08-838-399-1 Sequence 1, Appli
C 40 15.6 60.0 2481 3 US-09-235-839-1 Sequence 1, Appli
C 41 15.6 60.0 2481 4 US-09-327-035-1 Sequence 1, Appli
C 42 15.6 60.0 2604 2 US-08-630-118A-3 Sequence 3, Appli
C 43 15.6 60.0 2604 2 US-08-838-399-3 Sequence 3, Appli
C 44 15.6 60.0 2604 3 US-09-235-839-3 Sequence 3, Appli
C 45 15.6 60.0 2604 4 US-09-327-035-3 Sequence 3, Appli

```

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 100.0%; Score 26; DB 3; Length 4403765;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTCGGAGCCGACTACTTCTGGGT 26

Db 567850 GATTCGGAGCCGACTACTTCTGGGT 567825

##### RESULT 2

US-09-103-840A-1/c

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      100.0%; Score 26; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTCGGAGCCCGACTACTTCTGGGT 26
Db 566408 GATTCGGAGCCCGACTACTTCTGGGT 566383

RESULT 3
US-09-063-893A-2
; Sequence 2, Application US/09063893A
; Patent No. 6020464
; GENERAL INFORMATION:
; APPLICANT: Kenji OKAZAKI et al.
; TITLE OF INVENTION: A NOVEL SIGNAL TRANSDUCTION FACTOR AND A
; TITLE OF INVENTION: GENE ENCODING THE SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,893A
; FILING DATE: April 22, 1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; STRAIN: Swiss-Webster/NIH
US-09-063-893A-2

Query Match      65.4%; Score 17; DB 3; Length 209;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ATTTCGGAGCCCGACTACTTCTGGGT 26
Db 122 ACTCGGCGCCGACTTCTTCATGGT 146

us-10-624-714-8.rni

RESULT 4
US-09-560-198A-3
; Sequence 3, Application US/09560198A
; Patent No. 6492497
; GENERAL INFORMATION:
; APPLICANT: Thompson, Julia E
; APPLICANT: Lennard, Simon N
; APPLICANT: Wilton, Alison J
; APPLICANT: Braddock, Peta SH
; APPLICANT: Du Fou, Sarah L
; APPLICANT: McCafferty, John G
; APPLICANT: Conroy, Louise A
; APPLICANT: Tempest, Philip R
; TITLE OF INVENTION: Specific binding members for TGFbeta1
; FILE REFERENCE: 28111/35620A
; CURRENT APPLICATION NUMBER: US/09/560,198A
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/131,983
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-560-198A-3

Query Match      65.4%; Score 17; DB 4; Length 369;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTCGGAGCCCGACTACTTCTGGG 25
Db 316 GATACGAGCCCGACTACTTCTGGG 340

RESULT 5
US-09-487-444-3
; Sequence 3, Application US/09487444
; Patent No. 6159697
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF SMAD7 EXPRESSION
; FILE REFERENCE: RTS-0133
; CURRENT APPLICATION NUMBER: US/09/487,444
; CURRENT FILING DATE: 2000-01-19
; NUMBER OF SEQ ID NOS: 49
; SEQ ID NO 3
; LENGTH: 3111
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)...(1576)
US-09-487-444-3

Query Match      65.4%; Score 17; DB 3; Length 3111;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ATTTCGGAGCCCGACTACTTCTGGGT 26
Db 208 ACTCGGCGCCGACTTCTTCATGGT 232

RESULT 6
US-08-818-112-24
; Sequence 24, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

```

APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-24

Query Match 64.6%; Score 16.8; DB 3; Length 726;  
Best Local Similarity 90.0%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CGGAGCCCGACTACTTCTGG 24  
Db 268 CGGTGCCCGACTACTG 287

RESULT 7  
US-08-818-111-24  
Sequence 24, Application US/08818111  
Patent No. 633852  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
TUBERCULOSIS  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,111  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C6  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-111-24

Query Match 64.6%; Score 16.8; DB 4; Length 726;  
Best Local Similarity 90.0%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 CGGAGCCCGACTACTTCTGG 24  
Db 268 CGGTGCCCGACTACTG 287

RESULT 8  
US-09-056-556-24  
Sequence 24, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-056-556-24

Query Match 64.6%; Score 16.8; DB 4; Length 726;  
Best Local Similarity 90.0%; Pred. No. 72;

TREATY



QY 1 GATTCGGAGCCGACTACTTCTGGGT 26  
|||||  
Db 292 GATTCGGAGCTCAAGTACCTCTGGT 317

## RESULT 12

US-08-592-874-1/c  
; Sequence 1, Application US/08592874  
; Patent No. 5854034  
; GENERAL INFORMATION:  
; APPLICANT: POLLOCK, THOMAS J.  
; APPLICANT: YAMAZAKI, MOTOHIDE  
; APPLICANT: THORNE, LINDA  
; APPLICANT: MIKOLAJCZAK, MARCIA  
; APPLICANT: ARMENTROUT, RICHARD W.  
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING  
; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JULES E. GOLDBERG  
; STREET: 261 MADISON AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016-2391  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,874  
; FILING DATE:

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/377,440  
; FILING DATE: 24-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GOLDBERG, JULES E.  
; REGISTRATION NUMBER: 24,408  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-986-4090  
; TELEFAX: 212-818-9479  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28804 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FRAGMENT TYPE: N-terminal  
US-08-592-874-1

Query Match 63.1%; Score 16.4; DB 2; Length 28804;  
Best Local Similarity 76.9%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTTCTGGGT 26  
|||||  
Db 9638 GATTCGCTGCCGCTATCTCTGGCT 9613

## RESULT 13

US-09-096-942-2/c  
; Sequence 2, Application US/09096942  
; Patent No. 6027925  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Thomas J  
; APPLICANT: Mikolajczak, Marcia  
; APPLICANT: Yamazaki, Motohide  
; APPLICANT: Thorne, Linda  
; APPLICANT: Armentrout, Richard W  
; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria

; TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris  
; FILE REFERENCE: seq list for appl filed from pro. appl  
; CURRENT APPLICATION NUMBER: US/09/096,942  
; CURRENT FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/049,428  
; EARLIER FILING DATE: 1997-06-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; TYPE: DNA  
; LENGTH: 28804  
; ORGANISM: Sphingomonas sp. S88  
US-09-096-942-2

Query Match 63.1%; Score 16.4; DB 3; Length 28804;  
Best Local Similarity 76.9%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTTCTGGGT 26  
|||||  
Db 9638 GATTCGCTGCCGCTATCTCTGGCT 9613

## RESULT 14

US-09-096-867-2/c  
; Sequence 2, Application US/09096867  
; Patent No. 6030817  
; GENERAL INFORMATION:  
; APPLICANT: Pollock, Thomas J  
; APPLICANT: Mikolajczak, Marcia  
; APPLICANT: Yamazaki, Motohide  
; APPLICANT: Thorne, Linda  
; APPLICANT: Armentrout, Richard W  
; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria  
; FILE REFERENCE: seq list for appl filed from pro. appl  
; CURRENT APPLICATION NUMBER: US/09/096,867  
; CURRENT FILING DATE: 1998-06-11  
; EARLIER APPLICATION NUMBER: 60/049,428  
; EARLIER FILING DATE: 1997-06-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; TYPE: DNA  
; LENGTH: 28804  
; ORGANISM: Sphingomonas sp. S88  
US-09-096-867-2

Query Match 63.1%; Score 16.4; DB 3; Length 28804;  
Best Local Similarity 76.9%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGACTACTTCTGGGT 26  
|||||  
Db 9638 GATTCGCTGCCGCTATCTCTGGCT 9613

## RESULT 15

US-09-252-991A-10247  
; Sequence 10247, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

```

; SEQ ID NO 10247
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10247

Query Match      61.5%; Score 16; DB 4; Length 516;
Best Local Similarity 79.2%; Pred. NO. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      3 TTGCGAGCCGCGACTTCTCTGGGT 26
         |||||  |||||  |||||  |||||
DB      426 TTGCGAGCCGCGACTTCTCTGGGT 449

Search completed: April 29, 2004, 11:46:30
Job time : 56.8824 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 485.333 Seconds  
(without alignments)  
242.328 Million cell updates/sec

Title: US-10-624-714-8

Perfect score: 26

Sequence: 1 gattcgagcccgactacttctgggt 26

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US05\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	26	100.0	86114	15	US-10-080-170-648
C 2	18.2	70.0	531	15	US-10-156-761-5870
C 3	18.2	70.0	9025608	15	US-10-156-761-1
C 4	18	69.2	21581	15	US-10-017-161-2041
C 5	18	69.2	21581	16	US-10-232-798-1687
C 6	17.6	67.7	868	13	US-10-424-599-122323
C 7	17.6	67.7	1611	16	US-10-369-493-39638
C 8	17.6	67.7	1611	16	US-10-369-493-40002
C 9	17.6	67.7	3432	16	US-10-369-493-31453
C 10	17.6	67.7	5227	9	US-09-764-877-3458
C 11	17.6	67.7	5227	16	US-10-242-515-3458
C 12	17.6	67.7	119596	15	US-10-270-336-3
C 13	17.2	66.2	399	9	US-09-783-590-1530
C 14	17.2	66.2	1029	13	US-10-107-431-64

```

15 17.2 66.2 1209 15 US-10-156-761-3272 Sequence 3272, Ap
16 17.2 66.2 2136 16 US-10-104-047-618 Sequence 618, App
17 17.2 66.2 45055 13 US-10-107-431-277 Sequence 277, App
18 17.2 66.2 9025608 15 US-10-156-761-1 Sequence 1, Appli
19 17 65.4 325 9 US-09-960-352-315 Sequence 315, App
20 17 65.4 369 13 US-10-243-308-3 Sequence 3, Appli
21 17 65.4 369 15 US-10-243-268-3 Sequence 3, Appli
22 17 65.4 384 10 US-09-764-891-1265 Sequence 1265, Ap
23 17 65.4 1629 15 US-10-156-761-2710 Sequence 2710, Ap
24 17 65.4 2441 13 US-10-425-114-34390 Sequence 34390, A
25 17 65.4 3111 9 US-09-954-456-1877 Sequence 1877, Ap
26 17 65.4 3111 10 US-09-960-706-464 Sequence 464, App
27 17 65.4 3111 10 US-09-873-319-284 Sequence 284, App
28 17 65.4 3111 15 US-10-181-603-3 Sequence 3, Appli
29 17 65.4 3275 13 US-10-424-599-84682 Sequence 84682, A
30 17 65.4 3695 15 US-10-213-948-7 Sequence 7, Appli
31 17 65.4 3753 12 US-09-764-891-7624 Sequence 7624, Ap
32 17 65.4 69350 10 US-10-041-018-379 Sequence 379, App
33 16.8 64.6 726 15 US-10-193-002-24 Sequence 24, Appl
34 16.8 64.6 726 15 US-10-084-843-24 Sequence 24, Appl
35 16.6 63.8 253 13 US-10-424-599-57700 Sequence 57700, A
36 16.6 63.8 419 13 US-10-424-599-99680 Sequence 99680, A
37 16.6 63.8 565 13 US-10-424-599-131879 Sequence 131879, A
38 16.6 63.8 927 13 US-10-282-122A-33359 Sequence 33359, A
39 16.6 63.8 1014 16 US-10-369-493-42606 Sequence 42606, A
40 16.6 63.8 1937 15 US-10-157-669-16 Sequence 16, Appl
41 16.4 63.1 300 9 US-09-854-124-16 Sequence 16, Appl
42 16.4 63.1 305 13 US-10-424-599-84322 Sequence 84322, A
43 16.4 63.1 384 9 US-09-770-791-106 Sequence 106, App
44 16.4 63.1 395 13 US-10-424-599-41241 Sequence 41241, A
45 16.4 63.1 422 10 US-09-918-995-5652 Sequence 5652, Ap

```

#### ALIGNMENTS

##### RESULT 1

```

US-10-080-170-648/c
; Sequence 648, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; PRIOR FILING DATE: 2002-06-10
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 648
; LENGTH: 86114
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-648

```

```

Query Match 100.0%; Score 26; DB 15; Length 86114;
Best Local Similarity 100.0%; Pred No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 GATTCGAGCCCGACTACTTCTGGGT 26

Db 53007 GATTCGAGCCCGACTACTTCTGGGT 52982

##### RESULT 2

```

US-10-156-761-5870/c
; Sequence 5870, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI

```

```

; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5870
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(531)
; US-10-156-761-5870

Query Match          70.0%; Score 18.2; DB 15; Length 531;
Best Local Similarity 87.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGAGCCCGACTTCTGGGT 26
DB 347 TCGGAGCCGACTTGTCTGGGT 325

RESULT 3
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Query Match          70.0%; Score 18.2; DB 15; Length 9025608;
Best Local Similarity 87.0%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGAGCCCGACTTCTGGGT 26
DB 7102901 TCGGAGCCGACTTGTCTGGGT 7102923

us-10-017-161-2041
; Sequence 2041, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2041
; LENGTH: 21581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(21581)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(304)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11453)..(12511)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15724)..(16034)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21347)..(21381)
; US-10-017-161-2041

Query Match          69.2%; Score 18; DB 15; Length 21581;
Best Local Similarity 80.8%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTGGAGCCCGACTTCTGGGT 26
DB 647 GATTGGAGTCTGGCTCTCTGGGT 672

RESULT 5
US-10-292-798-1687
; Sequence 1687, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1687
; LENGTH: 21581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
```

```
; LOCATION: (1)...(21581)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(304)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11453)..(12511)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15724)..(16034)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21347)..(21381)
US-10-292-798-1687

Query Match          69.2%; Score 18; DB 16; Length 21581;
Best Local Similarity 80.8%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGCGACTACTTCTGGGT 26
Db 647 GATTCGGAGTCTGGTCTCTCTGGGT 672

RESULT 6
US-10-424-599-122323/c
; Sequence 122323, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 122323
; LENGTH: 868
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_81463C.1
US-10-424-599-122323

Query Match          67.7%; Score 17.6; DB 13; Length 868;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TTCGGAGCCGCGACTACTTCTGGGT 26
Db 827 TTCGGAGCGTGAATATTTCTGGAT 804

RESULT 7
US-10-369-493-39638
; Sequence 39638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31453
; LENGTH: 3432
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31453
```

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39638
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39638

Query Match          67.7%; Score 17.6; DB 16; Length 1611;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGCGACTACTTCTGG 24
Db 943 GATTCGGAGCGGCGATTACCTGTGG 966

RESULT 8
US-10-369-493-40002
; Sequence 40002, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40002
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-40002

Query Match          67.7%; Score 17.6; DB 16; Length 1611;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTCGGAGCCGCGACTACTTCTGG 24
Db 943 GATTCGGAGCGGCGATTACCTGTGG 966

RESULT 9
US-10-369-493-31453/c
; Sequence 31453, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31453
; LENGTH: 3432
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31453
```

Query Match 67.7%; Score 17.6; DB 16; Length 3432;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GATTCGAGCCGCGACTACTCTGG 24  
|||||  
Db 2441 GATTCGAGCCGCGACTACTCTGG 2418  
RESULT 10  
US-09-764-877-3458  
; Sequence 3458, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3458  
; LENGTH: 5227  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-3458

Query Match 67.7%; Score 17.6; DB 9; Length 5227;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ATTCGGAGCCGCGACTACTCTGG 25  
|||||  
Db 1819 AGTCGAGCCTCCTCTCTGG 1842

RESULT 11  
US-10-242-515-3458  
; Sequence 3458, Application US/10242515  
; Publication No. US20040009488A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005C1  
; CURRENT APPLICATION NUMBER: US/10/242,515  
; CURRENT FILING DATE: 2002-09-13  
; PRIOR APPLICATION NUMBER: 09/764,877  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3458  
; LENGTH: 5227  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-10-242-515-3458

Query Match 67.7%; Score 17.6; DB 16; Length 5227;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ATTCGGAGCCGCGACTACTCTGG 25  
|||||  
Db 1819 AGTCGAGCCTCCTCTCTGG 1842

RESULT 12  
US-10-270-336-3/c  
; Sequence 3, Application US/10270336  
; Publication No. US20030074678A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHU, Shiaooping et al.  
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL001146CON  
; CURRENT APPLICATION NUMBER: US/10/270,336  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 60/270,873  
; PRIOR FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 119596  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(119596)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-270-336-3

Query Match 67.7%; Score 17.6; DB 15; Length 119596;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GATTCGAGCCGCGACTACTCTGG 24  
|||||  
Db 63223 GATTCGAGCCTGAGTACTCTGG 63200

RESULT 13  
US-09-783-590-1530/c  
; Sequence 1530, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 15.2  
; FILE REFERENCE: FO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1530  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature

LOCATION: (9)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (11)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (69)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (123)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (203)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (257)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (274)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (325)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (350)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (370)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (388)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (389)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (391)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (392)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-1530

Query Match 66.2%; Score 17.2; DB 9; Length 399;  
Best Local Similarity 82.6%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ATTCGAGCCCGACTACTTCTGG 24  
DB 362 ATTCGAGCCCGACTACTTCTGG 340

RESULT 14  
US-10-107-431-64  
Sequence 64, Application US/10107431  
Publication No. US20030224364A1  
GENERAL INFORMATION:  
APPLICANT: Staffa, Chris  
APPLICANT: Zazopoulos, Emanuel  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
FILE REFERENCE: 3001-7US  
CURRENT APPLICATION NUMBER: US/10/107,431  
CURRENT FILING DATE: 2002-03-28  
NUMBER OF SEQ ID NOS: 282  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 64  
LENGTH: 1029  
TYPE: DNA  
ORGANISM: Streptomyces mobaraensis  
US-10-107-431-64

Query Match 66.2%; Score 17.2; DB 13; Length 1029;  
Best Local Similarity 86.4%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGAGCCCGACTACTTCTGG 25  
DB 592 TCGGAGCCCGACTACTTCTGG 613

RESULT 15  
US-10-156-761-3272  
Sequence 3272, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 3272  
LENGTH: 1209  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1209)  
US-10-156-761-3272

Query Match 66.2%; Score 17.2; DB 15; Length 1209;  
Best Local Similarity 86.4%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGGAGCCCGACTACTTCTGGT 26  
DB 1109 CGGAGCCCGACTACTTCTGGT 1130

Search completed: April 29, 2004, 20:44:57  
Job time : 501.333 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 29, 2004, 02:23:09 ; Search time 198.118 Seconds  
(without alignments)  
514.627 Million cell updates/sec

Title: US-10-624-714-9  
Perfect score: 24  
Sequence: 1 cgcgcagagccgatgccgaagc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	24	100.0	110000	AAI99682_11	Continuation (12 o
2	24	100.0	110000	AAI99683_11	Continuation (12 o
3	19.8	82.5	3564	ABL16858	Abi21699 Drosophil
4	19.8	82.5	6607	ABL21698	Abi21698 Drosophil
5	19.8	82.5	58957	AA58471	Aa58471 Nucleotid
6	19.4	80.8	42000	AA63349	Aa63349 Streptom
7	19.4	80.8	63164	AA63348	Aa63348 Streptom
8	19.2	80.0	345	ABQ91564	Abq91564 M. capsul
9	19.2	80.0	526	ABZ73106	Abz73106 Rice leaf
10	19.2	80.0	653	ACG60116	Acc60116 Rice endo
11	19.2	80.0	1560	ADA70449	Ada70449 Rice gene
12	18.8	78.3	629	AAH26306	Aah26306 Maize roo
13	18.8	78.3	629	AAH76444	Aah76444 DNA encod
14	18.8	78.3	987	AAH26307	Aah26307 Maize roo
15	18.8	78.3	987	AAH26308	Aah26308 Maize roo
16	18.8	78.3	987	AAH76447	Aah76447 DNA encod
17	18.8	78.3	3496	AAQ14465	Aaql14465 Maltopen
18	18.8	78.3	4834	ABL18279	Abi18279 Drosophil
19	18.8	78.3	7983	ABL18278	Abi18278 Drosophil
20	18.8	78.3	68750	AA255887	Aa255887 Sorangium
21	18.8	78.3	71989	AAA29349	Aaa29349 Sorangium
22	18.4	76.7	236	AAI15806	Aaxi15806 cDNA enco
23	18.4	76.7	376	ABL16859	Abi16859 Drosophil

24	18.4	76.7	488	2	AAI15804	Aaxi15804 cDNA enco
25	18.4	76.7	894	7	ACA37564	Prokaryot
26	18.4	76.7	2376	4	ABL16858	Drosophil
27	18.2	75.8	261	6	ABL73022	Corn tass
28	18.2	75.8	270	8	ADA49152	Adaa49152 Maize gen
29	18.2	75.8	272	6	ABL74661	Abi74661 Corn tass
30	18.2	75.8	463	3	AA434395	Aac434395 Zea mays
31	18.2	75.8	579	6	ABQ91849	Abq91849 M. capsul
32	18.2	75.8	579	6	ABQ91850	Abq91850 M. capsul
33	18.2	75.8	708	6	ABQ53438	Abq53438 Oligonuc
34	18.2	75.8	708	6	ABQ53439	Abq53439 Oligonuc
35	18.2	75.8	741	9	ADB80222	Adb80222 Mycobacte
36	18.2	75.8	780	6	ABU51218	Abi51218 Thermotol
37	18.2	75.8	780	6	ABN80345	Abn80345 B. caldot
38	18.2	75.8	795	7	ACA38665	Aca38665 Prokaryot
39	18.2	75.8	798	7	ACA40702	Aca40702 Prokaryot
40	18.2	75.8	1047	7	ADA69761	Ada69761 Rice gene
41	18.2	75.8	1134	7	ACA3753	Aca3753 Prokaryot
42	18.2	75.8	1146	5	AAF61289	Aaf61289 N. magada
43	18.2	75.8	1149	7	ACA23742	Aca23742 Prokaryot
44	18.2	75.8	2442	4	AAS51448	Aas51448 Pseudomon
45	18.2	75.8	2442	7	ACA19443	Aca19443 Prokaryot

ALIGNMENTS

RESULT 1  
AAI99682\_11  
Continuation (12 of 45) of AAI99682 from base 1100001 (Mycobacterium tuberculosis strain: LOCUS AAI99682 Accession Aai99682

WP	Sequence	split into	45 fragments	LOCUS	AAI99682	Accession	Aai99682
WP	Fragment Name	Begin	End				
WP	AAI99682_00	1	110000				
WP	AAI99682_01	100001	210000				
WP	AAI99682_02	200001	310000				
WP	AAI99682_03	300001	410000				
WP	AAI99682_04	400001	510000				
WP	AAI99682_05	500001	610000				
WP	AAI99682_06	600001	710000				
WP	AAI99682_07	700001	810000				
WP	AAI99682_08	800001	910000				
WP	AAI99682_09	900001	1010000				
WP	AAI99682_10	1000001	1110000				
WP	AAI99682_11	1100001	1210000				
WP	AAI99682_12	1200001	1310000				
WP	AAI99682_13	1300001	1410000				
WP	AAI99682_14	1400001	1510000				
WP	AAI99682_15	1500001	1610000				
WP	AAI99682_16	1600001	1710000				
WP	AAI99682_17	1700001	1810000				
WP	AAI99682_18	1800001	1910000				
WP	AAI99682_19	1900001	2010000				
WP	AAI99682_20	2000001	2110000				
WP	AAI99682_21	2100001	2210000				
WP	AAI99682_22	2200001	2310000				
WP	AAI99682_23	2300001	2410000				
WP	AAI99682_24	2400001	2510000				
WP	AAI99682_25	2500001	2610000				
WP	AAI99682_26	2600001	2710000				
WP	AAI99682_27	2700001	2810000				
WP	AAI99682_28	2800001	2910000				
WP	AAI99682_29	2900001	3010000				
WP	AAI99682_30	3000001	3110000				
WP	AAI99682_31	3100001	3210000				
WP	AAI99682_32	3200001	3310000				
WP	AAI99682_33	3300001	3410000				
WP	AAI99682_34	3400001	3510000				
WP	AAI99682_35	3500001	3610000				
WP	AAI99682_36	3600001	3710000				
WP	AAI99682_37	3700001	3810000				
WP	AAI99682_38	3800001	3910000				
WP	AAI99682_39	3900001	4010000				
WP	AAI99682_40	4000001	4110000				

WP AAI99682\_41 4100001 4210000  
WP AAI99682\_42 4200001 4310000  
WP AAI99682\_43 4300001 4410000  
WP AAI99682\_44 4400001 4411529

Query Match 100.0%; Score 24; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24  
|||  
Db 12852 CGCCGACGAGCGCGATGCCGAAGC 12875

RESULT 2  
AAI99683\_11  
Continuation (12 of 44) of AAI99683 from base 1100001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

Fragment Name	Begin	End
WP AAI99683_00	1	110000
WP AAI99683_01	100001	210000
WP AAI99683_02	200001	310000
WP AAI99683_03	300001	410000
WP AAI99683_04	400001	510000
WP AAI99683_05	500001	610000
WP AAI99683_06	600001	710000
WP AAI99683_07	700001	810000
WP AAI99683_08	800001	910000
WP AAI99683_09	900001	1010000
WP AAI99683_10	1000001	1110000
WP AAI99683_11	1100001	1210000
WP AAI99683_12	1200001	1310000
WP AAI99683_13	1300001	1410000
WP AAI99683_14	1400001	1510000
WP AAI99683_15	1500001	1610000
WP AAI99683_16	1600001	1710000
WP AAI99683_17	1700001	1810000
WP AAI99683_18	1800001	1910000
WP AAI99683_19	1900001	2010000
WP AAI99683_20	2000001	2110000
WP AAI99683_21	2100001	2210000
WP AAI99683_22	2200001	2310000
WP AAI99683_23	2300001	2410000
WP AAI99683_24	2400001	2510000
WP AAI99683_25	2500001	2610000
WP AAI99683_26	2600001	2710000
WP AAI99683_27	2700001	2810000
WP AAI99683_28	2800001	2910000
WP AAI99683_29	2900001	3010000
WP AAI99683_30	3000001	3110000
WP AAI99683_31	3100001	3210000
WP AAI99683_32	3200001	3310000
WP AAI99683_33	3300001	3410000
WP AAI99683_34	3400001	3510000
WP AAI99683_35	3500001	3610000
WP AAI99683_36	3600001	3710000
WP AAI99683_37	3700001	3810000
WP AAI99683_38	3800001	3910000
WP AAI99683_39	3900001	4010000
WP AAI99683_40	4000001	4110000
WP AAI99683_41	4100001	4210000
WP AAI99683_42	4200001	4310000
WP AAI99683_43	4300001	4403765

Query Match 100.0%; Score 24; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24  
|||  
Db 12875 CGCCGACGAGCGCGATGCCGAAGC 12898

RESULT 3  
ABL21699  
ID ABL21699 standard; DNA; 3564 BP.

XX ABL21699;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16570.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ds.

OS Drosophila melanogaster.

FN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Claim 1; SEQ ID NO 16570; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-ABBS72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3564 BP; 865 A; 1106 C; 1058 G; 535 T; 0 U; 0 Other;

Query Match 82.5%; Score 19.8; DB 4; Length 3564;

Best Local Similarity 91.3%; Pred. No. 1.4e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAG 23  
|||

Db 744 CGACGACGAGCGCGATGCCGAAG 766

RESULT 4

ABL21698/c

ID ABL21698 standard; DNA; 6607 BP.

XX ABL21698;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16567.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ds.

OS Drosophila melanogaster.



```

PN WO200171042-A2.
XX 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 16567; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ: Sequence 6607 BP; 1558 A; 1589 C; 1586 G; 1874 T; 0 U; 0 Other;
Query Match 82.5%; Score 19.8; DB 4; Length 6607;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 CGCCGACGAGCCGATGCCGAAG 23
DB 4258 CGACGACGAGCCGATGCCGAGG 4236

RESULT 5
AAA58471/c
ID AAA58471 standard; DNA; 58857 BP.
XX
XX AC AAA58471;
XX
XX DT 31-OCT-2000 (first entry)
XX
XX DE Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
XX BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
XX OS Streptomyces verticillus.
XX
XX FH Key Location/Qualifiers
XX CDS 223..564
XX /tag= a
XX /transl_except= (pos: 1..3, aa: Met)
XX /note= "ORF 30; encodes AAB07556"
XX
XX CDS 561..2309
XX /tag= b
XX /transl_except= (pos: 1..3, aa: Met)
XX /note= "ORF 29; encodes AAB07557"
XX
XX CDS 2767..3486
XX /tag= c
XX /note= "ORF 28; encodes AAB07558"
XX
XX CDS 3527..5593
XX /tag= d
XX

```

WO200040704-A1.

13-JUL-2000.

PF	06-JAN-2000; 2000WO-US000445.	FT	CDS	complement(3442..4971)
XX		FT		/*tag= e
PR	06-JAN-1999; 99US-0115435P.	FT		/product= "ORF -3 protein"
PR	05-FEB-1999; 99US-0118848P.	FT		5982..7479
PR	05-JAN-2000; 2000US-00477962.	FT		/*tag= f
XX		FT		/product= "glycerol phosphate transporter"
PA	(REGC ) UNIV CALIFORNIA.	FT		complement(7573..9900)
XX		FT		/*tag= g
XX	Shen B, Du L, Sanchez C, Chen M, Edwards DU;	FT		/product= "ABC transport/Uvra-like protein"
XX		FT		complement(9982..11349)
XX		FT		/*tag= h
DR	WPI; 2000-465974/40.	FT		/product= "Na+/H+ transporter"
DR	P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,	FT		complement(11351..12835)
DR	AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,	FT		/*tag= i
DR	AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,	FT		/product= "hydroxylase/halogenase"
DR	AAB07576, AAB07577, AAB07578.	FT		13012..14079
XX		FT		/*tag= j
FT	New bleomycin gene cluster components useful for peptide and/or	FT		/product= "dNDP-glucose synthase"
PT	polyketide metabolites, especially bleomycin, production and for	FT		complement(14212..14643)
PT	chemically modifying biological molecules.	FT		/*tag= k
XX		FT		/product= "CagA"
PS	Claim 8; Page 97-136; 162pp; English.	FT		complement(14690..15922)
XX		FT		/*tag= l
CC	The present sequence represents the BLM (Bleomycin) gene cluster,	FT		/product= "aminotransferase"
CC	containing open reading frames (ORFs) 8-30. The proteins encoded by the	FT		complement(15919..16653)
CC	gene cluster are useful for producing peptides and/or polyketide	FT		/*tag= m
CC	metabolites, especially bleomycin or bleomycin analogues. They are also	FT		/product= "N-methyl transferase"
CC	useful for chemically modifying biological molecules to produce branched	FT		complement(16653..17924)
CC	methyl groups, and for coupling amino acids and fatty acids. They may be	FT		/*tag= n
CC	reacted with an apo-carrier protein and coenzyme A to produce a holo-	FT		/product= "C-methyl transferase"
CC	carrier protein. The BLM gene cluster or catalytic domains can be used	FT		complement(18031..19191)
CC	individually or collectively to produce thiazolidine, thiazoline,	FT		/*tag= o
CC	bithiazoline and bithiazoline-containing microbial metabolites. The BLM	FT		/product= "spoxide hydrazase"
CC	gene cluster may also be used to produce sugars	FT		complement(19267..19929)
XX		FT		/*tag= p
SQ	Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;	FT		/product= "anthranilate synthase II"
	Query Match 82.5%; Score 19.8; DB 3; Length 58857;	FT		complement(19926..21407)
	Best Local Similarity 91.3%; Pred. No. 1.3e+02;	FT		/*tag= q
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	FT		/product= "anthranilate synthase I"
		FT		complement(21424..22878)
QY	2 GCCGACGAGCGCGATGCCGAGC 24	FT		/*tag= r
		FT		/product= "coenzyme F390 synthetase"
		FT		complement(22875..23546)
Db	50206 GCCGTCGAGCGCGACGCCGAGC 50184	FT		/*tag= s
		FT		/product= "iron-sulphur flavoprotein"
		FT		complement(23566..24702)
RESULT 6		FT		/*tag= t
AAA63349/c		FT		/product= "O-acyl transferase"
ID AAA63349 standard; DNA; 42000 BP.		FT		complement(24986..25564)
XX		FT		/*tag= u
AC	AAA63349;	FT		/product= "epimerase"
XX		FT		25815..27170
DT	06-MAR-2001 (first entry)	FT		/*tag= v
DE		FT		/product= "monooxygenase"
DE		FT		27214..28593
XX		FT		/*tag= w
XX	Streptomyces globisporus C-1027 gene cluster ORF -7-25.	FT		/product= "glycosyl transferase"
KW	Erediyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;	FT		28590..29588
KM	cancer; ds.	FT		/*tag= x
XX		FT		/product= "dNDP glucose dehydratase"
OS	Streptomyces globisporus.	FT		29632..31197
XX		FT		/*tag= y
XX		FT		/product= "SgcB transmembrane efflux protein"
FT		FT		31280..32590
FT		FT		/*tag= z
FT		FT		/product= "coenzyme F390 synthetase"
FT		FT		32809..34392
FT		FT		/*tag= aa
FT		FT		/product= "hydroxylase"
FT		FT		complement(34458..35294)
FT		FT		/*tag= ab
FT		FT		/product= "ORF 22 protein"
FT		FT		/transl_except= (pos:35223..35225;aa:Ile)

FT CDS /transl\_except= (pos:35226. .35228,aa:Ala)  
FT complement(35518. .35938)  
FT /\*tag= ac  
FT /product= "ORF 18 protein"  
FT 36165. .37490  
FT /\*tag= ad  
FT /product= "coenzyme F390 synthetase"  
FT 37559. .38938  
FT /\*tag= ae  
FT /product= "type II NRPS condensation enzyme"  
FT 38983. .39264  
FT /\*tag= af  
FT /product= "type II peptidyl carrier protein"  
FT complement(39367. .40986)  
FT /\*tag= ag  
FT /product= "aminomutase"  
FT complement(41052. .42000)  
FT /\*tag= ah  
FT /product= "type II NRPS adenylation enzyme"  
FT /partial  
XX WO200040596-A1.  
XX  
XX 13-JUL-2000.  
XX  
XX 06-JAN-2000; 2000WO-US000446.  
XX  
XX 06-JAN-1999; 99US-0115434P.  
PR 05-JAN-2000; 2000US-00477962.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Shen B, Liu W, Christenson SD, Standage S;  
XX  
XX WPI: 2000-465947/40.  
DR P-FSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,  
DR AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,  
DR AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,  
DR AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,  
DR AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587.  
XX  
XX Isolated nucleic acid comprising a nucleic acid encoding any of C-1027  
PT open reading frames (ORFs) -7 to 42, excluding ORF 9 (cagA), useful for  
PT the production of enediyne C-1027 antitumor antibiotics.  
XX  
XX Claim 1; Page 78-130; 160pp; English.  
XX  
XX The present sequence is the first 42000 bases of the enediyne C-1027 gene  
CC cluster from Streptomyces globisporus. Enediyne C-1027 is an antibiotic,  
CC consisting of an apoprotein and a non-peptidic chromophore, which acts by  
CC damaging DNA. The sequences within the gene cluster, and the proteins  
CC they encode, can be used in the treatment of cancer, along with  
CC antagonists of the protein. Each of the open reading frames is  
CC specifically claimed, excluding ORF 9, which encodes CagA  
XX  
XX Sequence 42000 BP; 6282 A; 14213 C; 15007 G; 6498 T; 0 U; 0 Other;  
SQ  
Query Match 80.8%; Score 19.4; DB 3; Length 42000;  
Best Local Similarity 95.2%; Pred. No. 1.8e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGCCGACGAGCGCGATGCCGA 21  
Db 2249 CGCCGACGAGCGCGAGCCGA 2229  
RESULT 7  
AAA63348/c  
ID AAA63348 standard; DNA; 63164 BP.  
XX  
XX AAA63348;  
XX  
DT 06-MAR-2001 (first entry)

XX Streptomyces globisporus C-1027 gene cluster.  
DE  
XX  
XX Enediyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;  
XX cancer; ds.  
XX  
XX Streptomyces globisporus.  
XX  
XX Key Location/Qualifiers  
CDS complement(8. .658)  
FT /\*tag= a  
FT /product= "ORF -7 protein"  
FT complement(930. .1478)  
FT /\*tag= b  
FT /product= "ORF -6 protein"  
FT complement(1649. .2713)  
FT /\*tag= c  
FT /product= "ORF -5 protein"  
FT complement(2850. .3237)  
FT /\*tag= d  
FT /product= "ORF -4 protein"  
FT complement(3442. .4971)  
FT /\*tag= e  
FT /product= "ORF -3 protein"  
FT 5982. .7479  
FT /\*tag= f  
FT /product= "glycerol phosphate transporter"  
FT complement(7573. .9900)  
FT /\*tag= g  
FT /product= "ABC transport/UvrA-like protein"  
FT complement(9982. .11349)  
FT /\*tag= h  
FT /product= "Na+/H+ transporter"  
FT complement(11351. .12835)  
FT /\*tag= i  
FT /product= "hydroxylase/halogenase"  
FT 13012. .14079  
FT /\*tag= j  
FT /product= "dNDP-glucose synthase"  
FT complement(14212. .14643)  
FT /\*tag= k  
FT /product= "CagA"  
FT complement(14690. .15922)  
FT /\*tag= l  
FT /product= "aminotransferase"  
FT complement(15919. .16653)  
FT /\*tag= m  
FT /product= "N-methyl transferase"  
FT complement(16653. .17924)  
FT /\*tag= n  
FT /product= "C-methyl transferase"  
FT complement(18031. .19191)  
FT /\*tag= o  
FT /product= "epoxide hydrazide"  
FT complement(19267. .19929)  
FT /\*tag= p  
FT /product= "anthranilate synthase II"  
FT complement(19926. .21407)  
FT /\*tag= q  
FT /product= "anthranilate synthase I"  
FT complement(21424. .22878)  
FT /\*tag= r  
FT /product= "coenzyme F390 synthetase"  
FT complement(22875. .23546)  
FT /\*tag= s  
FT /product= "iron-sulphur flavoprotein"  
FT complement(23566. .24702)  
FT /\*tag= t  
FT /product= "O-acyl transferase"  
FT complement(24986. .25564)  
FT /\*tag= u  
FT /product= "epimerase"  
FT 25815. .27170

```

FT FT      /tag= v      complement="monooxygenase"
FT FT      /product= "monooxygenase"
FT FT      27214..28593
FT FT      /tag= w      /product= "glycosyl transferase"
FT FT      28590..29588
FT FT      /tag= x      /product= "dNDP glucose dehydratase"
FT FT      29632..31197
FT FT      /tag= y      /product= "SgcB transmembrane efflux protein"
FT FT      31280..32590
FT FT      /tag= z      /product= "coenzyme F390 synthetase"
FT FT      32809..34392
FT FT      /tag= aa     /product= "hydroxylase"
FT FT      complement(34458..35294)
FT FT      /tag= ab     /product= "ORF 22 protein"
FT FT      /transl_except= (pos:35223..35225,aa:Ile)
FT FT      /transl_except= (pos:35226..35228,aa:Ala)
FT FT      complement(35518..35938)
FT FT      /tag= ac     /product= "ORF 18 protein"
FT FT      36165..37490
FT FT      /tag= ad     /product= "coenzyme F390 synthetase"
FT FT      37559..38938
FT FT      /tag= ae     /product= "type II NRPS condensation enzyme"
FT FT      38983..39284
FT FT      /tag= af     /product= "type II peptidyl carrier protein"
FT FT      complement(39367..40986)
FT FT      /tag= ag     /product= "aminomutase"
FT FT      /product= "aminomutase"
FT FT      complement(41052..42611)
FT FT      /tag= ah     /product= "type II NRPS adenylation enzyme"
FT FT      43945..46024
FT FT      /tag= ai     /product= "transmembrane transport protein"
FT FT      46167..47171
FT FT      /tag= aj     /product= "O-methyl transferase"
FT FT      47228..48484
FT FT      /tag= ak     /product= "P450 hydroxylase"
FT FT      complement(48607..49714)
FT FT      /tag= al     /product= "oxidoreductase"
FT FT      50350..51390
FT FT      /tag= am     /product= "ORF 31 protein"
FT FT      complement(51421..52341)
FT FT      /tag= an     /product= "oxidoreductase"
FT FT      53241..54074
FT FT      /tag= ao     /product= "ORF 33 protein"
FT FT      54231..55379
FT FT      /tag= ap     /product= "ORF 34 protein"
FT FT      56026..56880
FT FT      /tag= aq     /product= "proline oxidase"
FT FT      complement(56927..57736)
FT FT      /tag= ar     /product= "ORF 36 protein"
FT FT      complement(57833..58304)
FT FT      /tag= as     /product= "ORF 37 protein"
FT FT      /product= "ORF 37 protein"

CDS      complement(58440..60095)
FT      /tag= at
FT      /product= "P450 hydroxylase"
CDS      complement(60092..60621)
FT      /tag= au
FT      /product= "ORF 39 protein"
CDS      60940..62019
FT      /tag= av
FT      /product= "ORF 40 protein"
CDS      62045..62899
FT      /tag= aw
FT      /product= "ORF 41 protein"
CDS      complement(62787..63164)
FT      /tag= ax
FT      /product= "ORF 42 protein"
XX      WO200040596-A1.
XX      13-JUL-2000.
XX      06-JAN-2000; 2000WO-US000446.
XX      06-JAN-1999; 99US-0115434P.
XX      03-JAN-2000; 2000US-00477962.
XX      (REGC ) UNIV CALIFORNIA.
XX      Shen B, Liu W, Christenson SD, Standage S;
XX      WPI; 2000-465947/40.
XX      P-PSDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,
XX      AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,
XX      AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,
XX      AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,
XX      AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587,
XX      AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594,
XX      AAB13595, AAB13596, AAB13597, AAB13598, AAB13600, AAB13601, AAB13602,
XX      AAB13603, AAB13606, AAB13607.
XX      Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX      open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for
XX      the production of enediyne C-1027 antitumor antibiotics.
XX      Claim 1; Page 78-157; 160pp; English.
XX      The present sequence is the enediyne C-1027 gene cluster from
XX      Streptomyces globisporus. Enediyne C-1027 is an antibiotic, consisting of
XX      an apoprotein and a non-peptidic chromophore, which acts by damaging DNA.
XX      The sequences within the gene cluster, and the proteins they encode, can
XX      Query Match      80.8%; Score 19.4; DB 3; Length 63164;
XX      Best Local Similarity 95.2%; Pred. No. 1.8e+02;
XX      Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX      QY      1 CGCCGACGAGGCGCGATGCCGA 21
XX      DB      2249 CGCCGACGAGGCGCGAGCGCA 2299
XX      RESULT 8
XX      ABQ91564
XX      ID      ABQ91564 standard; DNA; 345 BP.
XX      AC      ABQ91564;
XX      DT      01-OCT-2002 (first entry)
XX      DE      M. capsulatus gene #1549 for DNA array.
XX      KW      Micro array; gene; ds; differential expression; gene expression.
XX      OS      Methylococcus capsulatus.
XX

```



```
XX SQ Sequence 653 BP; 126 A; 209 C; 189 G; 129 T; 0 U; 0 Other;
Query Match 80.0%; Score 19.2; DB 7; Length 653;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24
Db 390 CGCCGACGAGCGCGATGCCGAGC 413

RESULT 11
ADA70449/c
ID ADA70449 standard; DNA; 1560 BP.
XX AC ADA70449;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3772.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX KM gene, ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX SY (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WI; 2003-175290/17.
XX Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX Claim 6; SEQ ID NO 3772; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
XX involved in plant resistance or response to pathogenic infection. M1
XX comprises identifying a gene whose expression is significantly altered in
XX the incompatible interaction of plant gene expression relative to
XX expression of the gene in an uninfected plant, in a mutant plant that
XX does not express a gene associated with response to pathogenic infection,
XX or in a corresponding incompatible or compatible interaction. (M1) is
XX useful for conferring resistance to resistance or tolerance to a plant to
XX bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.
XX SQ Sequence 1560 BP; 232 A; 550 C; 483 G; 295 T; 0 U; 0 Other;
Query Match 80.0%; Score 19.2; DB 7; Length 1560;
Best Local Similarity 87.5%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24
Db 628 CGCCGACGAGCGCGATGCCGAGC 605

RESULT 12
AAH26306/c
XX ID ADA70444 standard; DNA; 629 BP.
XX AC AAH76444;

AAH26306 standard; cDNA; 629 BP.
XX AC AAH26306;
XX DT 02-OCT-2001 (first entry)
XX DE Maize root transcriptional factor cDNA.
XX KW Root transcriptional factor; maize; corn; transgenic plant;
XX KW crop improvement; drought tolerance; ss.
XX OS Zea mays.
XX FH Key Location/Qualifiers
XX FT CDS 1..579
XX FT /*tag= a
XX PN WO200157201-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US002920.
XX PR 01-FEB-2000; 2000US-0178916P.
XX PR 19-JAN-2001; 2001US-00766112.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Bruce WB;
XX WI; 2001-483434/52.
XX P-PSDB; AAB82614.
XX Isolated nucleic acid, useful for modulating expression in maize,
XX soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton rice, barley
XX or millet plants.
XX Claim 1(a); Page 53-54; 59pp; English.
XX The present sequence is that of cDNA encoding a maize root
XX transcriptional factor. The cDNA can be amplified from a maize nucleic
XX acid library using primers that selectively hybridize to loci within the
XX sequence. The invention provides isolated root transcriptional factor
XX nucleic acids and their encoded proteins, as well as methods and
XX compositions relating to altering root transcriptional factor levels in
XX plants. Recombinant expression cassettes comprising a maize root
XX transcription factor polynucleotide operably linked, in sense or
XX antisense orientation, to a promoter are claimed. Also claimed are host
XX cells, and transgenic soybean, sunflower, sorghum, canola, wheat,
XX alfalfa, cotton, rice, barley, millet and (especially) maize plants
XX comprising the recombinant expression cassettes. The polynucleotides and
XX encoded proteins can be expressed temporally or spatially, e.g. at
XX developmental stages, in tissues and/or in quantities which are
XX characteristic of the non-transgenic plants. The plants show increased
XX lateral root initiation and growth typical of hys mutants, which could
XX enhance root anchorage and/or drought tolerance
XX SQ Sequence 629 BP; 140 A; 182 C; 253 G; 54 T; 0 U; 0 Other;
Query Match 78.3%; Score 18.8; DB 4; Length 629;
Best Local Similarity 90.9%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CGCAGCAGCGCGATGCCGAGC 24
Db 187 CGCAGCAGCGCGATGCCGAGC 166

RESULT 13
AAH76444/c
XX ID AAH76444 standard; DNA; 629 BP.
XX AC AAH76444;
```

XX 22-OCT-2001 (first entry)  
XX DNA encoding maize root transcriptional factor #1.  
XX  
XX  
XX Maize; growth stimulant; root transcriptional factor; HYS stimulator;  
XX antisense therapy; gene expression control; root initiation; root growth;  
XX ds.  
XX  
XX Zea mays.  
XX  
XX Key Location/Qualifiers  
XX CDS 1..579  
XX /\*tag= a  
XX /product= "root transcriptional factor"  
XX  
XX WO200157200-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 22-JAN-2001; 2001WO-US002147.  
XX  
XX 01-FEB-2000; 2000US-0178916P.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Bruce WB;  
XX WPI; 2001-465704/50.  
XX P-PSDB; AAG66525.  
XX An isolated protein encoding HYS useful for modulating the level of root  
XX transcriptional factor in a plant especially monocots for stimulating  
XX growth.  
XX  
XX Claim 1; Page 53-54; 56pp; English.  
XX The invention relates to an isolated protein comprising a polypeptide  
XX comprising at least 20 contiguous amino acids of, or at least 80%  
XX identical to, two 192 amino acid sequences fully defined in the  
XX specification. The polynucleotide encoding the polypeptide is useful for  
XX modulating the level of root transcriptional factor in a plant. The  
XX polypeptide and the polynucleotide are useful for controlling the  
XX expression of genes involved in root initiation and growth, including  
XX responses to environmental or pathogenic cues. The polynucleotide is  
XX useful for designing probes or amplification primers in the detection,  
XX quantification, or isolation of gene transcripts. The polypeptide can be  
XX used to obtain antibodies. The present sequence encodes one of the two  
XX 192 amino acid sequences from Zea mays which is provided in the  
XX specification  
XX  
XX Query Match 78.3%; Score 18.8; DB 5; Length 629;  
XX Best Local Similarity 90.9%; Pred. No. 3.6e+02;  
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 3 CCGACGAGCCGATCGGAAGC 24  
XX 187 CCGACGAGCCGACGCGGACGC 166  
XX  
XX  
XX RESULT 14  
XX AAH26307/c  
XX ID AAH26307 standard; cDNA; 987 BP.  
XX  
XX AC AAH26307;  
XX  
XX 02-OCT-2001. (first entry)  
XX  
XX Maize root transcriptional factor cDNA.  
XX  
XX Root transcriptional factor; maize; corn; transgenic plant;  
XX Root transcriptional factor; maize; corn; transgenic plant;

KW crop improvement; drought tolerance; ss.  
XX Zea mays.  
XX  
XX Key Location/Qualifiers  
XX CDS 93..671  
XX /\*tag= a  
XX  
XX WO200157201-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US002920.  
XX  
XX 01-FEB-2000; 2000US-0178916P.  
XX 19-JAN-2001; 2001US-00766112.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Bruce WB;  
XX WPI; 2001-483434/52.  
XX P-PSDB; AAB82615.  
XX Isolated nucleic acid, useful for modulating expression in maize,  
XX soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton rice, barley  
XX or millet plants.  
XX Claim 1(a); Page 55-56; 59pp; English.  
XX The present sequence is that of cDNA encoding a maize root  
XX transcriptional factor. The cDNA can be amplified from a maize nucleic  
XX acid library using primers that selectively hybridise to loci within the  
XX sequence. The invention provides isolated root transcriptional factor  
XX nucleic acids and their encoded proteins, as well as methods and  
XX compositions relating to altering root transcriptional factor levels in  
XX plants. Recombinant expression cassettes comprising a maize root  
XX transcription factor polynucleotide operably linked, in sense or  
XX antisense orientation, to a promoter are claimed. Also claimed are host  
XX cells, and transgenic soybean, sunflower, sorghum, canola, wheat,  
XX alfalfa, cotton, rice, barley, millet and (especially) maize plants  
XX comprising the recombinant expression cassettes. The polynucleotides and  
XX encoded proteins can be expressed temporally or spatially, e.g. at  
XX developmental stages, in tissues and/or in quantities which are  
XX uncharacteristic of the non-transgenic plants. The plants show increased  
XX lateral root initiation and growth typical of hys mutants, which could  
XX enhance root anchorage and/or drought tolerance  
XX  
XX Sequence 987 BP; 195 A; 274 C; 357 G; 161 T; 0 U; 0 Other;  
XX  
XX Query Match 78.3%; Score 18.8; DB 4; Length 987;  
XX Best Local Similarity 90.9%; Pred. No. 3.6e+02;  
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX 3 CCGACGAGCCGATCGGAAGC 24  
XX 279 CCGACGAGCCGACGCGGACGC 258  
XX  
XX  
XX RESULT 15  
XX AAH26308/c  
XX ID AAH26308 standard; cDNA; 987 BP.  
XX  
XX AC AAH26308;  
XX  
XX 02-OCT-2001 (first entry)  
XX  
XX Maize root transcriptional factor cDNA.  
XX  
XX Root transcriptional factor; maize; corn; transgenic plant;  
XX crop improvement; drought tolerance; ss.  
XX Zea mays.  
XX

XX Key Location/Qualifiers  
FH 93.671  
FT /\*tag= a  
FT  
XX  
PN WC200157201-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US002920.  
XX  
XX 01-FEB-2000; 2000US-0178916P.  
PR 19-JAN-2001; 2001US-00766112.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Bruce WB;  
PI  
XX  
XX WPI; 2001-483434/52.  
DR P-PSDB; AAB82616.  
DR  
XX  
XX Isolated nucleic acid, useful for modulating expression in maize,  
PT soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton rice, barley  
PT or millet plants.  
PT  
XX  
PS Claim 1(a); Page 57-58; 59pp; English.  
XX  
XX The present sequence is that of cDNA encoding a maize root  
CC transcriptional factor. The cDNA can be amplified from a maize nucleic  
CC acid library using primers that selectively hybridise to loci within the  
CC sequence. The invention provides isolated root transcriptional factor  
CC nucleic acids and their encoded proteins, as well as methods and  
CC compositions relating to altering root transcriptional factor levels in  
CC plants. Recombinant expression cassettes comprising a maize root  
CC transcription factor polynucleotide operably linked, in sense or  
CC antisense orientation, to a promoter are claimed. Also claimed are host  
CC cells, and transgenic soybean, sunflower, sorghum, canola, wheat,  
CC alfalfa, cotton, rice, barley, millet and (especially) maize plants  
CC comprising the recombinant expression cassettes. The polynucleotides and  
CC encoded proteins can be expressed temporally or spatially, e.g. at  
CC developmental stages, in tissues and/or in quantities which are  
CC uncharacteristic of the non-transgenic plants. The plants show increased  
CC lateral root initiation and growth typical of hy5 mutants, which could  
CC enhance root anchorage and/or drought tolerance  
XX  
SQ Sequence 987 BP; 194 A; 274 C; 358 G; 161 T; 0 U; 0 Other;

Query Match 78.3%; Score 18.8; DB 4; Length 987;  
Best Local Similarity 90.9%; Pred. No. 3.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCGACGAGCGCGATCGCGAGC 24  
|||  
Db 279 CCGACGAGCGCGAGCGCGAGC 258

Search completed: April 29, 2004, 04:57:12  
Job time : 201.218 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 362.165 Seconds  
(without alignments)  
2872.264 Million cell updates/sec

Title: US-10-624-714-9

Perfect score: 24

Sequence: 1 cgcgcagcaggccgatccgaagc 24

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pri.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rtd.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vrt.\*  
38: em\_sy.\*  
39: em\_hgo\_hum.\*  
40: em\_hgo\_mus.\*  
41: em\_hgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	24	100.0	20916	1	AE006986	AE006986 Mycobacte
2	24	100.0	327650	1	BX248337	BX248337 Mycobacte
3	24	100.0	349306	15	BX842575	BX842575 Mycobacte
C 4	20.8	86.7	2751	8	AK111663	AK111663 Oryza sat
C 5	20.8	86.7	5989	8	OS4427582	AJ427982 Oryza sat
C 6	20.8	86.7	10322	1	AE005135	AE005135 Halobacte
C 7	20.8	86.7	149437	2	AP005610	AP005610 Oryza sat
C 8	20.8	86.7	166679	2	AP005192	AP005192 Oryza sat
10	20.8	86.7	178073	1	SC0590464	AL590464 Streptomy
11	20.8	86.7	273785	1	SM5591793	AL591793 Sinorhizo
12	20.4	85.0	11773	1	AE012243	AE012243 Xanthomon
C 13	20.4	83.3	346547	1	AP003012	AP003012 Mesorhizo
C 14	19.8	82.5	1194	6	AR386658	AR386658 Sequence
15	19.8	82.5	5027	3	BT009960	BT009960 Drosophil
16	19.8	82.5	7686	1	SC0414671	AJ414671 Streptomy
17	19.8	82.5	14625	1	AE007264	AE007264 Sinorhizo
C 18	19.8	82.5	31317	1	SCSCP2	AL645771 Streptomy
C 19	19.8	82.5	40123	2	AC014130	AC014130 Drosophil
C 20	19.8	82.5	65961	2	AC016523	AC016523 Drosophil
C 21	19.8	82.5	77457	1	AF210249	AF210249 Streptomy
22	19.8	82.5	95626	8	AP005501	AP005501 Oryza sat
23	19.8	82.5	100589	8	OSJN00191	AL662991 Oryza sat
C 24	19.8	82.5	111230	8	AP005515	AP005515 Oryza sat
C 25	19.8	82.5	124366	8	AC131374	AC131374 Oryza sat
26	19.8	82.5	124420	8	OSJN00042	AL606597 Oryza sat
C 27	19.8	82.5	137981	8	AP004348	AP004348 Oryza sat
28	19.8	82.5	146951	8	AP003270	AP003270 Oryza sat
29	19.8	82.5	149948	2	AP005518	AP005518 Oryza sat
30	19.8	82.5	150465	8	AC091749	AC091749 Oryza sat
31	19.8	82.5	151490	8	AP003611	AP003611 Oryza sat
32	19.8	82.5	154696	8	AC093181	AC093181 Oryza sat
C 33	19.8	82.5	156232	8	AC108884	AC108884 Oryza sat
34	19.8	82.5	156551	2	AP005916	AP005916 Oryza sat
35	19.8	82.5	162005	2	AP005916	AP005916 Oryza sat
C 36	19.8	82.5	163307	3	AC011703	AC011703 Drosophil
C 37	19.8	82.5	165525	2	AP005837	AP005837 Oryza sat
C 38	19.8	82.5	167688	3	AC011702	AC011702 Drosophil
C 39	19.8	82.5	171437	8	AC135597	AC135597 Oryza sat
C 40	19.8	82.5	173281	3	AC012161	AC012161 Drosophil
C 41	19.8	82.5	179199	8	AC092548	AC092548 Oryza sat
C 42	19.8	82.5	183197	2	BX470161	BX470161 Danio rer
C 43	19.8	82.5	185194	2	AC133004	AC133004 Oryza sat
C 44	19.8	82.5	298900	1	AP005937	AP005937 Bradyrhiz
45	19.8	82.5	300029	8	AE017076	AE017076 Oryza sat

# ALIGNMENTS

RESULT 1  
AE006986  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

20916 bp DNA linear BCT 27-APR-2001  
Mycobacterium tuberculosis CDC1551, section 72 of 280 of the  
complete genome.  
AE006986 AE000516  
AE006986.1 GI:13880583  
Mycobacterium tuberculosis CDC1551  
Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 20916)  
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Pred. No. is the number of results predicted by chance to have a

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 20916)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
Location/Qualifiers  
1. .20916  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="Genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
complement(95. .1096)  
/gene="WT1018"  
complement(95. .1096)  
/gene="WT1018"  
/note="similar to SP:P55785 PID:1731693; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polyprenyl synthase"  
/protein\_id="AAK45265.1"  
/db\_xref="GI:13880584"  
/translation="MRMTATSMTPAVSLGDPQPTANVHDCIARITELINSELQAOBE VRDVAHLVDAGGPFPLFTVLAAQLGSDPDGMEVTVAGAAELMHLGTLCHDRV DESMSRTPTDNTWNNFALAGDYRFATASQASGLSDPEAFVAVEAFELITQ MEATGPASHDITIBHCLRVVHEKTGSLIAASGQALSGAEQIRRVARLGRMIG AAFELSRDIIAISGDSATLSGADLQAOVHTLPMELVAREQTPDTSRLRELLAGPIHD HVAEALTLRCSPGIGKAKNVAAVAAQAARELYLPDPRPRALATLIDHVASCD" complement(1133. .1789)  
/gene="WT1019"  
complement(1133. .1789)  
/gene="WT1019"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45266.1"  
/db\_xref="GI:13880585"  
/translation="MAESLNPISLVSRISAFRLPDWTRTVRARFPAAGLVMLAGVAA LRSNPEDRSVVAHDLRPTALTPTDVRLEKRSATLTPDGSQADLDAVVGSTLAS PTRKEVLTDVLLGSLRAESTAGPDARIVPLHADSALVDLVNVRGVDVLAAPVTD SPAALRLATDAIVLVLSAQKAQADSRLVVALPARLANTVAGALGQVTLTLH" complement(1862. .2215)  
/gene="WT1020"  
complement(1862. .2215)  
/gene="WT1020"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45267.1"  
/db\_xref="GI:13880586"  
/translation="MTFPEVLVPTYSYECTQANRFDVQVATDDALTTCERCSGRIL KLFNAGVVFVKGTFGYRTDSRESGKSKQTSNGSTSESTKSGSSGSSGSSKSG STEKSTSTTAAAV" complement(2268. .2855)  
/gene="WT1021"  
complement(2268. .2855)  
/gene="WT1021"  
/note="similar to GP:4455709; identified by sequence

similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="5-formyltetrahydrofolate cyclo-ligase-related protein"  
/protein\_id="AAK45268.1"  
/db\_xref="GI:13880587"  
/translation="MASKALRDQLLAARRRADDVRAAEARMLRGHLRMVTSSTV CAVPVGSGPSIEMLDVLRAGRVLLPVARTAGDPLPLRWGEYRAGGLARWGL LPPPPWLPALQAQSLVLPALAVDQGVRLGRGFRGYDRSLRCRDPHARLVAVV RTVELVDVLPSEPHDVPMTHTALTPERGLIAPGGE"  
2962. .3882  
/gene="WT1022"  
2962. .3882  
/gene="WT1022"  
/note="similar to GB:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UTP-glucose-1-phosphate uridylyltransferase"  
/protein\_id="AAK45269.1"  
/db\_xref="GI:13880588"  
/translation="MSRPBVLTPFTAIVPAAGLGRFLPATKTVPKELLPVVDTPGIE LVAAEAAAAGAERLVITVSEKDGVAHFVEDLVLEGTLEARGKAMLAKEARAPALI KYESVQARPLGLHAIGVEPTLSPEDAVALLPDDLVLPTGLEMTSKVRSRGG TVLCAIEVAREBESISANGVDFVPEPDGYTDDPNVLKVRGMVEKEKAETAPSRVAAAG RYVLDRAFDALRRIDRAGGVEQLTDAITALLAIEGHPVHVVVHQSRRHDLGNPGYL KAAVDPALDRDDYGPDLRRLVRLGLTQ"  
3959. .5239  
/gene="WT1023"  
3959. .5239  
/gene="WT1023"  
/note="similar to GB:AE000666; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="molybdopterin biosynthesis protein MoeA"  
/protein\_id="AAK45270.1"  
/db\_xref="GI:13880589"  
/translation="MRSVEEQARISAAVAPRPPIRVAIAEAOGLMCAEEVVTPEMP GFDAIDGIAVRSVDVAGVGTGGVFADHGDLDGRDVLTLTPVNGTIEAGATLSR LQPRQAVRTGAPLPTLADAVLPLRWTDGMSRVRLRGPSPGAYVRAGDDVQPGD VAVRAQTIIGAAGLAAVGRVRLVHPRLSVYMGELVLSRTFGQGVYDVN SYLAAAGRDACAENRVGIVNSNDTELGEIVEGOLNRAEVVVIAGVGGAAEAVRS VLSLGEVMEVVRVAVHGPSVQGFGLGRDGVPTLLPANPVSAIVVFVWVRLRLS LCKRHMERIVSARTLSPTTSVAGKGYLGRQMDQSDGEVLVQALGAPGASSHLL ATLAEANCLVVPVPTGAEQIRTGELVDAFLAQHG"  
5302. .5913  
/gene="WT1024"  
5302. .5913  
/gene="WT1024"  
/note="similar to GB:M99278 SP:P09454 GB:X06118 PID:147655 PID:42744; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal-protein-alanine acetyltransferase"  
/protein\_id="AAK45271.1"  
/db\_xref="GI:13880590"  
/translation="NAVGLRVSAGVIRLPRVMDGVHWSKIRLADRAHLEPWEPSA DGEVTRHTVAWPAVCSGLRSEARNGRMLPVVIELDQGFQCGTLTGNVTHALRSAW IGWVPSATGGVATGALGLDCHFCFGLMHRVEATVRPENASRAVLAKVGFREE GLRRILEVDRAWRDHLMLAITVEVYGSVASTLVRAHASWP"  
5975. .7150  
/gene="WT1025"  
5975. .7150  
/gene="WT1025"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45272.1"  
/db\_xref="GI:13880591"  
/translation="MNYRCVIALGALTTPRPWFTMGPRGNRRREQVIMPSIQSILLMI



VAVVQAAALSELRRPDRPRAETNVEFWAAVVLDFARVPANMPAMFTCCRTAGWCA  
 HILEQRLGKLVPSAIYVGPGRSPESVDGWERVILTA"  
 complement (3413..6061)  
 /locus\_tag="Mb0914c"  
 CDS  
 complement (3413..6061)  
 /locus\_tag="Mb0914c"  
 /note="Mb0914c, -, len: 882 aa. Equivalent to RV0890c,  
 len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (99.8% identity in 882 aa overlap). Probable  
 transcriptional regulatory protein, LuxR family, highly  
 similar (but shorter 238 aa in N-terminus) to  
 NP\_302202.1|NC\_002677 possible transcriptional regulator  
 from Mycobacterium leprae (1106 aa). Also highly similar  
 (generally in part) to others e.g. T50568 probable  
 multi-domain regulatory protein from Streptomyces  
 coelicolor (1334 aa); P10957|NARL\_ECOLI nitrate/nitrite  
 response regulator protein from Escherichia coli (216 aa),  
 FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99  
 aa overlap); etc. Also highly similar to others from  
 Mycobacterium tuberculosis e.g. MTCY02B10.22, MTW008.44,  
 MIV036\_21, and MTC131\_24. Contains P900017 Atp/GTP-binding  
 site motif A (P-loop), P900622 Bacterial regulatory  
 proteins, luxR family signature, and probable helix-turn  
 helix motif from aa 836 to 857 (Score 1559, +4.50 SD).  
 BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL  
 REGULATORS."  
 /codon\_start=1  
 /transl\_table=11  
 /product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN  
 (PROBABLY LUXR-FAMILY)"  
 /protein\_id="CAD93775.1"  
 /db\_xref="GI:31617666"  
 /translation="MRALLAQNRLVTLCTGGVGVKTRLAIOIASASLRLDLCFVDLA  
 PTESGIVATARAAGVLPDQGRSTMSLRPFIGNRMLVLDNCEHLLDCAALVY  
 ELUGPEITLIRTEPIGAGEITWRVPSNITDEAVELFADRSRQVGFTHAN  
 LRAAGVEICRRDLGLDLEFAFAARVSRMSPLEIADGLDDCFRLLAGVAGVQRQOT  
 NAAIDWSHALITETEQILFRRLAPFVGGFDLAAVRAVAGSDLDGPFVLDQLTLVD  
 KSLVADDCQGTTRVLETVRRYALEKLGSGEADVHARRDYTTALAASLNPADN  
 DQRLVARAETIDNRLAFAFVRENGHITLALQASLSIPIWFGRAHLRGLGWFSN  
 ILEDOHFHLVSTAVARALADKAMLTWLATS PVGATDIIAPAQAALAMAREVGD  
 ALIARALACGSSGYNEAARAPFAETDLARAIDDKWTLICQLIWRGVTGSGDP  
 NALRAAEFCRLDALTDFRVSRHCSLWLSLAQWAGNLTLELSRETAERASIN  
 DVPTGLVGLYQAQVLA YCGASAAHAIACIAAATELGQVYGGIYAAMTYALAAAG  
 DYTAALEASDAARPIIRAQPDQVTHQVIMQAALAGGDAIARQFANDAYDANGWH  
 RVVALTIRAVATARGEPELARDDAHALACAGELHIYQGMPEMELIAGLAGEVGH  
 SEGVLLGAAALROOTROVRPKIWDAGVQASVTALREMGDEDPDRAWAEGAALST  
 EAIYAQRGEGRKRPAGWSGLTTERDVRLVNLSEGLSNKDIKRLVFSRVTQTHL  
 THYAKLGLASRVQLVDEARAGSFS"  
 complement (6063..6920)  
 /locus\_tag="Mb0915c"  
 complement (6063..6920)  
 /locus\_tag="Mb0915c"  
 /note="Mb0915c, -, len: 285 aa. Equivalent to RV0891c,  
 len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (99.8% identity in 285 aa overlap). Possible  
 transcriptional regulator, highly similar in N-terminus to  
 NP\_302202.1|NC\_002677 possible transcriptional regulator  
 from Mycobacterium leprae (1106 aa). Also highly similar  
 to several Mycobacterium tuberculosis putative  
 transcriptional regulators e.g. Q1102|MTCY02B10.22  
 PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),  
 FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in  
 247 aa overlap); MIV036\_21; MTW008.44; MTCY02B10\_23. Also  
 shows similarity with several adenylate cyclases and  
 hydrolases from other organisms."  
 /codon\_start=1  
 /transl\_table=11  
 /product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"  
 /protein\_id="CAD93776.1"  
 /db\_xref="GI:31617667"  
 /translation="WLFNAVNSLPNTIDHAILRGEDHHPPTCAKCVARGRISALGS  
 LDRYHSURCYAAPDGRCEFPVPRVLIANQGLDVSRLPPTGTVTLLEADYEEST  
 HLWQMCPEDMATAIHLDTHTSEAITNTHGVQPVKRYEGDSFVAAFTRASDAACALD

LQRTSLAPIRLRLGLHTGEVQLDELYVGPTINRTARLRLDLAHHGGQVLSAATGDLVT  
 GRIPADAWLVLDLGRHPLRPRPEWYMQLCHPDIREKFPPLRTAKSSPTSIILPAQFTT  
 FVGRRAQIS"  
 7318..8805  
 /locus\_tag="Mb0916"  
 CDS  
 7318..8805  
 /locus\_tag="Mb0916"  
 /EC\_number="1.14.-.-"  
 /note="Mb0916, -, len: 495 aa. Equivalent to RV0892, len:  
 495 aa, from Mycobacterium tuberculosis strain H37Rv,  
 (100.0% identity in 495 aa overlap). Probable  
 monooxygenase (EC 1.14.-.-), highly similar to others e.g.  
 NP\_250787.1|NC\_002516 probable flavin-binding  
 monooxygenase from Pseudomonas aeruginosa (491 aa);  
 CAS95688.1|ALU32674 monooxygenase from Streptomyces  
 coelicolor (519 aa); P12015|CYMO\_ACIS cyclohexanone  
 monooxygenase from Acinetobacter sp. (542 aa), FASTA  
 scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa  
 overlap); etc. Also highly similar to RV0565c, RV3854c,  
 RV3083, etc from Mycobacterium tuberculosis. Has  
 hydrophobic stretch at N-terminus."  
 /codon\_start=1  
 /transl\_table=11  
 /product="PROBABLE MONOOXYGENASE"  
 /protein\_id="CAD93777.1"  
 /db\_xref="GI:31617668"  
 /translation="MTGRCPVAVVAGAGSMGVATLLSAGITDVCIEKADYVGGT  
 WRNTYPTGLTCDVPSRLYQYSPAKNPNWQMSRGGEIQDYLRGIAERGLRHRRFG  
 AVTVSARFDDGRWLRTDSGTSTDFLSATGLVHLHPRIPTPIAGLDDFRGTVFHSR  
 WDTVPFLGRRIATIGTSTGLVGLAGVAGKVTQFTQWLRVLPNFRYSKLAR  
 VFHRAFPCLGSLKAYSAFETFAVSLNGLRKLGVAVCSLRVLRVDPRLRLAL  
 TPDEYPMCKRLVMSGGFYRAIORDDVDLGTAGIDHVRHGI VTDGVLHEVDIVLAT  
 GPDSHAFERPMOLTGRDGI RIDDVWDGPHAQTVAI PGFPNFMGLGPHSPVGNPPL  
 TAVAESQAEHIVQWIKRMRHGEPTMEPKSAEANTVLRAMPNTVVTTCDSWYL

Query Match 100.0%; Score 24; DB 1; Length 327650;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGCCGACGAGCCGATGCCGAGC 24  
 Db 126305 CGCCGACGAGCCGATGCCGAGC 126328  
 RESULT 3  
 BX842575  
 ID BX842575 standard; circular genomic DNA; PRO; 349306 BP.  
 AC BX842575; AL010186; AL021006; AL021897; AL021999; AL123456; Z92539;  
 AC Z93777; Z94752; Z95209; Z95210; Z95584; Z95585; Z98260;  
 SV BX842575.1  
 XX  
 XX 21-NOV-2003 (Rel. 77, Created)  
 XX 21-NOV-2003 (Rel. 77, Last updated, Version 1)  
 DE Mycobacterium tuberculosis H37Rv complete genome; segment 4/13  
 XX complete genome.  
 XX Mycobacterium tuberculosis H37Rv  
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
 OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.  
 XX [1]  
 RN MEDLINE; 98295987.  
 RX PUBMED; 9634230.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekai P., Badcock K.,  
 Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,  
 Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jajels K.,  
 Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,













gene NAGELHGMVDPGRRTGCGRDGHWEAYCSGNNIPYARLLADDAGVETALPLDSSG  
GFTAKDVECAADGDTAAHVVEQLGVNGIGVTLNVAQVAPLVVYVGGVALHNPEQ  
VLGPTRAYIQRFVSNPEIRLTTLGDDVLKGAIASALGTGTDSTHAP"

CDS  
/gene="VNG2631H"  
/complement(8549..9154)  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng2631h"  
/protein\_id="RA020665.1"  
/db\_xref="GI:10582010"  
/translation="MMEASVGDGMGGLLEVTLDAEEVLAAGALVDHTGDVRIDRSR  
AGAMRSVANTAREVEVTVRVVGGGAVRPAFDYPCETAAACELGAGGRVSVAMDG  
FVAAPGVSGVSGINAGTRMGMLTQFTGERTVYAGRGVRRVERVALAEGDSRVVA  
AANVAFEDRVVRAVERVSAMEDAASVGRFHGPGTVMVATR"

Query Match 86.7%; Score 20.8; DB 1; Length 10922;  
Best Local Similarity 91.7%; Pred. NO. 2.1e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 9402 CGCCGACGAGCGCGCGCGGAGC 9425

RESULT 7  
AP005610  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) chromosome 6 clone  
OSUNBA0032M14, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
ACCESSION  
AP005610  
VERSION  
AP005610.1 GI:22091082  
KEYWORDS  
HTG; HTGS PHASE2.  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS  
TITLE  
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC  
clone:OSUNBA0032M14  
Published Only in Database (2002)  
2 (bases 1 to 149437)  
Sasaki,T., Matsumoto,T. and Katayose,Y.  
Direct Submission

JOURNAL  
TITLE  
Submitted (01-AUG-2002) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7458)  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
Location/Qualifiers  
1..149437  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="6"  
/clone="OSUNBA0032M14"

ORIGIN

Query Match 86.7%; Score 20.8; DB 2; Length 149437;  
Best Local Similarity 91.7%; Pred. NO. 1.5e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 137143 CGCGACGAGCGCGATGCCGAGC 137166

RESULT 8  
AP005192/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) chromosome 6 clone P0485A07,  
\*\*\* SEQUENCING IN PROGRESS \*\*\*  
ACCESSION  
AP005192  
VERSION  
AP005192.1 GI:20975318  
KEYWORDS  
HTG; HTGS PHASE2.  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
AUTHORS  
TITLE  
Sasaki,T., Matsumoto,T. and Katayose,Y.  
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC  
clone:P0485A07  
Published Only in Database (2002)  
2 (bases 1 to 166679)  
Sasaki,T., Matsumoto,T. and Katayose,Y.  
Direct Submission

JOURNAL  
TITLE  
Submitted (15-MAY-2002) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7458)  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
Location/Qualifiers  
1..166679  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="6"  
/clone="P0485A07"

ORIGIN

Query Match 86.7%; Score 20.8; DB 2; Length 166679;  
Best Local Similarity 91.7%; Pred. NO. 1.4e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 61990 CGCGACGAGCGCGATGCCGAGC 61967

RESULT 9  
SC0590464  
LOCUS  
DEFINITION  
Streptomyces coelicolor plasmid SCF1; segment 2/2.  
ACCESSION  
AL590464  
VERSION  
AL590464.1 GI:13620666  
KEYWORDS  
4-hydroxy-2-oxovalerate aldolase; acetaldehyde dehydrogenase;  
AraC-family transcriptional regulator; bifunctional undecaprenyl  
pyrophosphate synthetase/phytoene synthase; bldA regulation; DNA

pyrophosphate synthetase/phytoene synthase; bldA regulation; DNA

integrase/recombinase; DNA-polymerase III, alpha chain;  
DNA-polymerase III, beta chain; DNA-primase/helicase; dnaE; dnaN;  
ECF-family sigma factor; eflux; esterase; helicase; hydratase;  
integral membrane; IS466S; lyase; methylenomycin A  
biosynthesis/resistance cluster; origin; plasmid partitioning  
protein, ParAB; plasmid transfer protein; sap; secreted;  
short-chain oxidoreductase; spore associated protein; Terminal  
inverted repeat; TerR; Tn4811; Tn5714; transposase; transposon.  
Streptomyces coelicolor A3(2)

## SOURCE

## ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

## REFERENCE

1 (bases 1 to 178073)

## AUTHORS

Redenbach,M., Ikeda,K., Yamasaki,M. and Kinashi,H.

## TITLE

Cloning and physical mapping of the EcoRI fragments of the giant

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

2 (bases 1 to 178073)

## AUTHORS

Brown,S.P., Murphy,L.D. and Harris,D.

## JOURNAL

## TITLE

Submitted (20-FEB-2001) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Plasmid supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded

by the BBSRC and Wellcome Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) The more  
significant matches with motifs in the PROSITE database are also  
included but some of these may be fortuitous. The length in codons  
is given for each CDS.

Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon. Gene  
prediction is based on positional base preference in codons using a  
specially developed Hidden Markov Model (Krogh et al., Nucleic  
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program  
of Bibb et al., Gene 30:157-66(1994).

## FEATURES

## source

1. 178073

/organism="Streptomyces coelicolor A3(2)"

/mol\_type="genomic DNA"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="plasmid SCP1"

200. .203

210. .488

210. .488

/gene="SCP1.184"

/gene="SCP1.184"

/note="SCP1.184, conserved hypothetical protein, len:

92aa; strongly similar to SCJ30.10c (TR:Q9SIX3

EMBL:AL109973) hypothetical protein from the chromosome

end(s) of Streptomyces coelicolor (92 aa) fasta scores;

opt: 499, z-score: 682.6, E(): 1.5e-30, 81.5% identity in

92 aa overlap. Also similar to SCB6.04 (TR:Q9KZT4

EMBL:AL353832) hypothetical protein from Streptomyces

coelicolor (75 aa) fasta scores; opt: 107, z-score: 161.9,

E(): 0.15, 30.3% identity in 81 aa overlap. Contains a TTA

encoded leucine at residue 3, possible target for bldA

regulation."

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

## RBS

## gene

## CDS

/protein\_id="CAC36706.1"  
/db\_xref="GI:13620667"  
/db\_xref="GOA:Q9ACW7"  
/db\_xref="SPTREMBL:Q9ACW7"  
/translation="VTLRFVGDIDPNTGGSGSPVVVEBSADLVQGEADELLKDLV  
GSENVAGKAGIPHERVIRIPAMVSVILREACNAERAAAEHRDVR"  
216. .218  
misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

/protein\_id="CAC36706.1"  
/db\_xref="GI:13620667"  
/db\_xref="GOA:Q9ACW7"  
/db\_xref="SPTREMBL:Q9ACW7"  
/translation="VTLRFVGDIDPNTGGSGSPVVVEBSADLVQGEADELLKDLV  
GSENVAGKAGIPHERVIRIPAMVSVILREACNAERAAAEHRDVR"  
216. .218  
misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

/protein\_id="CAC36706.1"  
/db\_xref="GI:13620667"  
/db\_xref="GOA:Q9ACW7"  
/db\_xref="SPTREMBL:Q9ACW7"  
/translation="VTLRFVGDIDPNTGGSGSPVVVEBSADLVQGEADELLKDLV  
GSENVAGKAGIPHERVIRIPAMVSVILREACNAERAAAEHRDVR"  
216. .218  
misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```
YLRSALMSTQPHSLSRVMSALRGSSEHSNLSLQTMALAMTASHVAHMEVEVLDPAS
PEDETSFGVDGVPAPSPGT"
complement (2547. .2831)
/gene="SCPI.188c"
complement (2547. .2831)
/gene="SCPI.188c"
/note="SCPI.188c, unknown, len: 94aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC36710.1"
/db_xref="GI:13620671"
/db_xref="GOA:O9ACW3"
/db_xref="SPTREMBL:Q9ACW3"
/translation="MGFWSNPDVLAVLANAPHIALLYVQAQCIREVGAIAMLIVLRG
TEASRPRIIAALGDLGRNKTVTIAQLRARRGRPKRYGPGASGKSGAG"
3133. .4272
/gene="SCPI.189"
3133. .4272
/gene="SCPI.189"
/note="SCPI.189, unknown, len: 379aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC36711.1"
/db_xref="GI:13620672"
/db_xref="GOA:O9ACW2"
/db_xref="SPTREMBL:Q9ACW2"
/translation="WAIGTISKILGGTYVPWSQLRLLERCGITDAPRLAQWQRIQ
DVRTHMRKEASNSRQLADDAHPLAVRFAQDLRGFLSTTQFTDDIEANTGSRVL
VMDISGLRALPLRWLVLLRASGEISQWLERRQNVAAEMKRAKAPYLVVVDPOL
PRYGSDSDYRPARMTMLDDWDLTWPILFOLPGDTSFDMVHLGVYFANLIDK
IAATELMTNIDRLARVCAQMTVNTGNEAARDPEAMCGEREDLLIARLARDFT
GVHCKQKSRIDSLRVTSATFQSVDRSPFGOSSDEAAAFVARLQGTTEVRFLE
TSAEAWSPDGRNSDLRPDDWSLDGRGTFISRIPTWLSKGR"
complement (4385. .5218)
/gene="SCPI.190c"
complement (4385. .5218)
/gene="SCPI.190c"
/note="SCPI.190c, possible integrase/recombinase, len:
277aa; similar to many eg. SW:P20184 (Yr23_STRFR)
hypothetical 23.1 kDa protein in transposon Tn4556 from
Streptomyces fradiae (205 aa) fasta scores; opt: 559,
z-score: 616.8, E(): 6.9e-27, 78.4% identity in 111 aa
overlap and TR:O30680 (EMBL:AF015088) putative plasmid
multimer resolution protein from plasmid pFAJ2600 of
Rhodococcus erythropolis (306 aa) fasta scores; opt: 276,
```

```
Query Match 86.7%; Score 20.8; DB 1; Length 178073;
Best Local Similarity 91.7%; Pred. No. 1.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCCAGCAGGCGCGATGCCGAGGC 24
|||||
Db 17690 CCGCCAGCAGGCGAGAGCGGAGGC 17713

RESULT 10
SME591793 273785 bp DNA linear BCT 05-JUL-2002
Sinorhizobium meliloti 1021 complete chromosome; segment 12/12.
ACCESSION AL591793 AL591688
VERSION AL591793.1 GI:15076142
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 273785)
Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boisnard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
```

```
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
21396507
PUBMED 11481430
REFERENCE 2 (bases 1 to 273785)
Gouzy,J.
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
Location/Qualifiers
FEATURES
1..273785
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
172..1461
/gene="SMC03265"
172..1461
/gene="SMC03265"
/EC_number="1.4.99.1"
/function="small molecule metabolism; degradation; amino
acids"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE AMINO ACID DEHYDROGENASE TRANSMEMBRANE
PROTEIN"
/protein_id="CAC47696.1"
/db_xref="GI:15076143"
/db_xref="GOA:Q92LE8"
/db_xref="SPTREMBL:Q92LE8"
/translation="MPEVSAIVVGAGIVGLSVAYGALRNGSVYIVDRNPEGDKASFG
NAGGIATVEVVPASVFLWKVDFWMLDPLGLALRMAHTPRLLPWLWNPARYGTASE
VRISTALADINGRVYDDLPLENEVGLAGDLLKGLATVYSENGLRNDMDHMDAKR
DLGVEQELTAASIREMEPALSPIVKNGVFTPOWSHKPKRVVDVLRHLVANGAEI
VRGEVRDVPTSDGAKVLDGGREMTARKIVIAAGAWSGRLAKQMGDRVLVSERGIT
TTIASPQGLALEREVIFADRFKVAITPLSIGLRIGGAAFFGLGVAHPYARSKALVKLAK
HYFFDIREGQVSGSRPTTDSLPFVIGRSGPSNNIIYAFGHGHLGLTQGTPTGKIV
SDLLGKRPKLIADIAPFSVARFSLKASVDFSDPLFLR"
1750..2041
/gene="f5 OR SMC03266"
3054..4070
/gene="SMC03267"
3054..4070
/gene="SMC03267"
/function="macromolecule metabolism; macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
```

```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE DIPEPTIDASE PROTEIN"
/protein_id="CAC47697.1"
/db_xref="GI:15076144"
/db_xref="GOA:Q92LE7"
/db_xref="SPTREMBL:Q92LE7"
/translation="MDALALHGDALFIDAVAPLLHKKOYGVSYVLGGVTCVAPTVGVG
SSPSTALKTNWLYRTNPGVRLVEKARHVEDAKDKGMGLFHFQGTGPTETDL
DLVADFAKVRVOLTNRKDYVGDGCEELDGLSFGKDLVRLNENRIIVDCSH
TGKTMDSLTTPAFVSHNSAVLPFSKRNIRDDQAKAANAATGVLAVTAPOEV
VSANGAARPSIDLIADHIDFVSVDGVLGDYDWMGLPDSFDGTGVVHEHHV
EAGDSQAPYPRPHYPPTPEITPELLKGTVGLIRRGYSEEDIRKILGLNWLRFHE
VWGE"
4070..5947
/gene="SMC03268"
CDS
4070..5947
/gene="SMC03268"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE PEPTIDE TRANSPORT SYSTEM ATP-BINDING
ABC TRANSPORTER PROTEIN"
/protein_id="CAC47698.1"
/db_xref="GI:15076145"
/db_xref="GOA:Q92LE6"
/db_xref="SPTREMBL:Q92LE6"
/translation="MARIEFISPLVDVNPKNYKSPVQVLSVNSCMSPRADGLWS
SVVQGVSPDIGARVAVVSGSGKSVTALSIMELNSPRGTKIEGVLGKGLLSL
SAEMROIRGNDIMFOEPMTSLNPVLTICFOIAESLILHRLGSLRAOENETIRLE
KVRIPAKRFREYHRSRSGNRQVMIAMALACPKLLIADPTTALDVTIQAILLE
LIKEQEDQMSVLFTHDMGVVADIRTVVMYQGVATVETADITDIFNRNHPYTRS
LISAVPKLSMAGRPRKRFVWVSTGEADVPEVPTVKSERPLEVAGLTKRFE
VKSQFSTVKGCVHVENYVSFIRAGETLALVSGSGCKSTTGRSVRLVFPSSGLVL
LDGVVYLKLSQNELRHQRNMOMIFQDPYALSNPRMNTGTAIBPLINGKQGSVA
DKVALLQVGLSPMAKRFPHFSGQORICIAARALAPKLIADVAEVSALDVSV
KSOVNLMLDLQARVGLAVLFIHDMVAVERVSVAVMVLGEIVAGSRAIQRNQC
HPTVRLLSAVPIADPARLQRTVSNDEIKSPFPADYVQVPRVPTGHTVMS
G"
6017..7615
/gene="SMC03269"
CDS
6017..7615
/gene="SMC03269"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE PEPTIDE-BINDING PERIPLASMIC ABC
TRANSPORTER PROTEIN"
/protein_id="CAC47699.1"
/db_xref="GI:15076146"
/db_xref="GOA:Q92LE5"
/db_xref="SPTREMBL:Q92LE5"
/translation="MTPATITIKTEIAALKLSTALVAGLLFTSANADTIRVMGNL
KIIDFIVTSGDTLTHAYMIDYKLFDEFESGVLPQLAEKVYDISADNKSFTIKRDLG
KFSDDGTFTSDVIOSLTWSARDQTLRLKIDNTLRIDLTSTFSPVPSA
LAGITCNPAIFMKRVASLPALTDTVSGSPYQDFSTWKPQGMRYTKSYIYPR
DBAASFYAGRKRAVMDRIEFSYVDPNTAMASMLTGKDIWLDPPMANAVARNPNL
KVYKVSQGLGVINHVPVPPDNVAKQALTYLIDQKEINALVAGDPDEFWACVAFLT

```

gene

CDS

```

CSGTGNESVYDGRGHPANVERKAKELFNEAGYDGOVVLDPDTPPEAHARALYLAQQ
LREIGKVLQAMDWCTLTSSRTSEKPAKGGNNLFPNVMGQOPASSPLTHLMJLASSC
EKAFWGPCDKQIEDLRASFRTVDIDKKEKIADELQARASAYLPFVMTGEYNNVMSY
GLNIEGFIPSAPEPPFPFVNRKXQ"
7638..8582
/gene="SMC03754"
CDS
7638..8582
/gene="SMC03754"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE PEPTIDE TRANSPORT SYSTEM PERMEASE ABC
TRANSPORTER PROTEIN"
/protein_id="CAC47700.1"
/db_xref="GI:15076147"
/db_xref="GOA:Q92LE4"
/db_xref="SPTREMBL:Q92LE4"
/translation="MIRYLIARILSAPVIGVTVIWFGLTYLPGDAASLAVGDYAS
PADIALRVKLHDDPIITQYLYWLNVLKDGHSIFSQPVLLALILSRVPTTLA
IAATLVAVVAVPLGLAAHNGGIADQVVSFGVAGTALPAFLGYLLIPAFGVYVG
SMPIAGFAGISGLEMAKHVLLPALMGFAQTALISRLIRASAFALMQDFVRVARA
KQDEITIMWHCLKPIAVFVITVGTSPFALMGVITETFAIPGIGSLLEVAIQH
KDPVITQGVLLLSLTVLVNLLVLSYLIVDFRIKYT"
8595..9437
/gene="SMC03272"
CDS
8595..9437
/gene="SMC03272"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology

```

Query Match 86.7%; Score 20.8; DB 1; Length 273785;  
 Best Local Similarity 91.7%; Pred. No. 1.4e+03;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24

Db 184474 CGCCGTCGAGCGGATGCCGAGC 184497

RESULT 11

AP005949

LOCUS

DEFINITION

AP005949

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

2

299850 bp DNA linear BCT 12-SEP-2003  
 Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 15/31.  
 AP005949 BA000040  
 AP005949.1 GI:27352050  
 Bradyrhizobium japonicum USDA 110  
 Bradyrhizobium japonicum USDA 110  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bradyrhizobaceae; Bradyrhizobium.  
 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,  
 Sasamoto, S., Watanabe, A., Igesawa, K., Iriguchi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimo, S., Teuruka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110  
 DNA Res. 9 (6), 189-197 (2002)  
 22484998  
 12597275





```
gene
2810..3364
/gene="XCC1433"
CDS
2810..3364
/gene="XCC1433"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA040730.1"
/db_xref="GI:21112499"
/translation="MSKTIKAVLVSLRAQSYNRQLAQAALAGDKATFEFLIGDL
PLYNDRSDYPABETRLKQRLRAADAVLFTPEYNSIFPGVLNAIDTASRPYCDSA
FAGKAAAGISVGAIGTALACQHLRGVLAIVHVLQCPPIYLQYKXGDLFGADGQIA
DADSRKFLQGFVDSPLAWIAHTR"
3553..4983
/gene="dnaB"
/note="synonym: XCC1434"
3553..4983
/gene="dnaB"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="replicative DNA helicase"
/protein_id="AA040731.1"
/db_xref="GI:21112500"
/translation="WSPARPGFSNRNRDRDRDDYDRPEPRDLQRLVRPHSVARCAVL
GGLMAPDFKIVKIDTENDPYRDHLYRAIRLSEKDRPDAVTLGMPESOCK
LEQVGDGAYLIELASTTSPANLAYAIVRDKAVLRQLIEVGNIVNDGFPQEGRES
VELLASAKVFAKAEAGARKTDFVAMPGALKDAFELNRNFRNGNITGLPTGYTD
FDAMTAGLOPDLIIILAAPAMGKTTLALNAYAAIKSKGVAVFMSMSASQLAMR
LISNGRINAGRLTGALDEEDWARVTGAIKMLKTKIFIDTTPGVSPVLRSKRRL
KRHDLGIVDIYQLMSVPGNSNRATEISEIRGLKGLAKELNVPVIALSOLNRL
ETRTDKRPMADRESGAIEODADMIVFYRDDYNNKSPDKGLAEIILIKHRRGGT
GSKLKFGEVTRFDNLADSVGSPE"
5238..6692
/gene="phr"
/note="synonym: XCC1435"
5238..6692
/gene="phr"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="photolase-like protein"
/protein_id="AA040732.1"
/db_xref="GI:21112501"
/translation="WLRGLTCRCVMSYATVWFRDLRLDNLRAALDAGHDPTP
LYIDAPHEGOWAPGASRAWHESLAALDASLRGSAALLIRQGDSDAQLDAVIAT
EAVAYWNRKYEPATQPRDAQIKSLRERGLEVQSCNALLFEFWILATQQGRPKVF
TPFWRNALTQRLPDAMPAPRSLPPLPASLDGVHDALNLLPTPAWDQGFWEHWQPSR
APAHMELEIFVDGSLGYSNRDRPDRVGTQSLPHLFGIEAPWRIASLEAQRSAR
NGADIDGVIROLGRDFAHYLLHFPDTTQNLNRPAGFDWATVDVPTLDAMQGRGT
GIPTVDAGLROLWTHWNRVMIVASLKLHVRHLEAGRWFTLVDADLANNT
MGWQWATGADAAPYRFVNPVQAQKFPQATYITRWIPELALPVKERFAPWLHP
LSLARLPTYPRAPIGLAEGRDAALAYAGTRG"
6826..7551
/gene="XCC1436"
6826..7551
/gene="XCC1436"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AA040733.1"
/db_xref="GI:21112502"
/translation="WSPAASKSLAVALASATLSACATGGSYVORDQVGDQTOQNET
GRNALITPAISVAGLITGDSATERRCHMIGAGIGALSGAVNGYQDRERALRENT
ANTGIDVORQDNTLNLPGDITDFGSKALKPQFYALNGVASTLREINQTVVEVVG
HTDSVGSVDVNRUSEERANAVAQYLTAAQGVORERMETMGAGKRYPIADNSTAGRAQ
```

```
NRREIRLIPROEGAASTTGMR"
complement(8305..9339)
/gene="XCC1437"
complement(8305..9339)
/note="XCC1437"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA040734.1"
/db_xref="GI:21112503"
/translation="MTHDLNDTLIFVKVVEQGSFIAANSGLPKTTVSRKQVELETR
LGRLLHRTTRRIGLTEAGAVHEHCORARELEAEASVQLOSGPRGWLRTVPYS
LGRTTWAPLLGEFHAQYBEIQLDHMHGNEKLDLIGGEADLAEVGLALPSNLVARKLG
SLRTQVSPSPSYIERGEPLHDELQFHETLALRKNRVNHRPFWSLSDGSDVRFEP
NPELVANDPAALNGAVLCGEGLLTGDMAPFFVQSGMVRVRLVAGTGFVDFNAVFF
AGRLVSPKVRAPFVDFELVRLNFDADYMAQCPARLAAQANGDAEVEVEAEALRAE
GGRIEKATA"
9648..10466
/gene="dauE"
/note="synonym: XCC1438; aknU"
9648..10466
/gene="dauE"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"

Query Match      85.0%; Score 20.4; DB 1; Length 11773;
Best Local Similarity 95.5%; Pred. No. 2.8e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCCGACGAGCCCGATCCGAAG 23
      |||||
Db 10919 GCCGACGAGCCCGATCCGAAG 10940

RESULT 13
LOCUS      AP003012/c
DEFINITION Mesorhizobium loti DNA, complete genome, section 19/21.
ACCESSION  AP003012 BA000012
VERSION     AP003012.2 GI:14026998
KEYWORDS
SOURCE      Mesorhizobium loti
ORGANISM    Mesorhizobium loti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1 (sites)
AUTHORS     Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
            Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
TITLE        DNA Res. 7 (6), 331-338 (2000)
JOURNAL      21082930
MEDLINE      11214968
PUBMED
REFERENCE    2 (bases 1 to 346547)
AUTHORS      Kaneko,T.
TITLE        Direct Submission
JOURNAL      Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research, Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
            On May 11, 2001 this sequence version replaced gi:11994987.
COMMENT
FEATURES
            source
                1..346547
                /organism="Mesorhizobium loti"
                /mol_type="genomic DNA"
                /strain="MAFF303099"
```



```
gene
/db_xref="taxon:381"
complement(4..507)
/gene="ml17507"
CDS
complement(4..507)
/gene="ml17507"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53953.1"
/db_xref="GI:14026993"
/transl_table="MPAFGGIPTWDFVRAVAIGIGATVLMIDLWAILLHKAFOQPP
NWPGVGVWHLGDVFFDDICEAPEYAEVALGWVHYFVGIVGILVVLACTGWL
RAPTELPAFLIGIVTGAGWELLAPGCMGAGWAAKRPNPNRIALNLVSHVTFVPGLGW
GTALLIR"
complement(574..1056)
/gene="ml17508"
CDS
complement(574..1056)
/gene="ml17508"
/codon_start=1
/transl_table=11
/product="peroxiredoxin-like protein"
/protein_id="BAB53954.1"
/db_xref="GI:14027000"
/transl_table="MTISVGDKLPEATKPTMTADGAKAITSABEIPGKKVVLFGVPGA
FTPTCSNHLPGYLNHDAILARGVDITAVSVVNDVHVMGAWARFTGGEDKILPLADG
SADFAKAVGLNDLSAGCGMLRSKRFSIMVDDGKVTALNIETKPGVDESAAHTLGL"
complement(1144..1950)
/gene="ml17509"
CDS
complement(1144..1950)
/gene="ml17509"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53955.1"
/db_xref="GI:14027001"
/transl_table="WQSLKILLGGAAGSASVLPASASSAWNSBEGKVLVTSGRP
DEAGRIQGVLDIALPGKWTYWRDPGDAGVPPQNLVNSASTNIADATISFPAPQHDHG
YKGAQYVPSVPLVFTLSAKOSADIVDFLIGICETICIPVQTRLTVDPGSDPN
ADDAALVAKSFAALPAPARSDFGINVLPQDHETLVVEASFPQDPEAADFVAGERDYM
FCAPRSKDGKLIITVILDRPSATPDGGLHYLTSSAGAVEGLLFPF"
2122..2730
/gene="ml17511"
CDS
2122..2730
/gene="ml17511"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53956.1"
/db_xref="GI:14027002"
/transl_table="WDLRHKRTAAGRCFLDDQFLIAMPGMKDRFRFTSVIYICAHSD
EGAMGLIINQTMFLPDLVLQGLIMNEQEARILPAQARDFVVRNGGFPVDRSGFVLH
SGDYRVSGLTVSDDICLTATVILRAISGGRGPRHMAWLGYSWGAGQLETBAEN
GWLTCAPSELLADIERKDYRILASIGIDLHLHLAAGHA"
complement(2752..4416)
/gene="ml17513"
CDS
complement(2752..4416)
/gene="ml17513"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53957.1"
/db_xref="GI:14027003"
/transl_table="WAGSDTVMDVLRDRVVTXPDVSIQGLAPNSLGAARNWLP
VLHADDRFTFTLVLEHRRGVAQNFRLGADGHVWFALRARPVIGSDGEVIR
VGTMDVTPQKSEERLHDVNDLGLPNRLEFMRLEAIISARTEEKVRPTVC
IDIDRFKQVNDLGI:SAGDTIILTIARLRLKPKDSLSRFAQDFALMLLSQDDPA
RIAGVADALKAINNPIITFAKREIVLTASIGLITWTSQTSADMVXDAELAMHQAR
FGDRLERPRAPVGTDRLOFESDLRRAERREFTLAYOPIVELEDGSVAGFEALL
RDHFRGRMIPADRTIPVASEGLVQLGLENMQQAEDLAGWOKIGDAPLSVSNL
SSQLIRLSDVRSVIRANLKPFCRLTESLNDNFEQTHVLTKLQGLIGL
SLDDFTGYSSLAYITRFPFDTKIDKSFVDDNTPKRAVLKLSMWNMAHELGLSVAAE
GISDRDALRELQMGCEYVQSPFMFGAPMGDPQVLTKLEQYPLTQA"
```

```
gene
complement(5644..6240)
/gene="ml17514"
CDS
complement(5644..6240)
/gene="ml17514"
/codon_start=1
/transl_table=11
/product="ribosomal-protein-alanine N-acetyltransferase"
/protein_id="BAB53958.1"
/db_xref="GI:14027004"
/transl_table="MPALPFFRRDLPAKGNLVTLRVPYTNDRWSTLRGESRAFLV
PEPMTPELDRTRAWRISRIREDIYAGGTATIAIFIEKSGKLAGIITGNIRHGV
SOSGHVGYWGERFCGRGLMTDAVKVARFATLRLHRIEAAACIPDNIRSIIVLEKA
GFRREGLSYLRINGIQWDHYLYARIADDDPLDAGTKD"
complement(6256..7548)
/gene="ml17516"
CDS
complement(6256..7548)
/gene="ml17516"
/codon_start=1
/transl_table=11
/product="processing proteinase"
/protein_id="BAB53959.1"
/db_xref="GI:14027005"
/transl_table="WQVEVSRSLNGLTVATETLPSIESVALGAWKSGARNERDEHG
MAHLEHMAFKTKRTAFETIASEIEDVGEINAASTVETTSYARVLSDDVPLAVDI
LADIQESFDFQELEREQHVILQIGAHDTPDDIVDFRTETAFRQTKGRSLGT
PDTKXFTSKQLHDFIERQYGAERVIIVAAAGDIKHDFVREYKQLGGRSKADSTIP
QYAVYGGDFREDRLMDAQIFGVFEGRAYHVRDFYASQVLSILGGMSSRLFOEVR
EKRGLCYSVYAFHWGFSDTGIFGVHAATGOSDIABLPVVIDELQKAGENILQEELDR
ARAQYPAGLIMSAPASPAISOIARQLLLFGPIAKEELMERLSALTVERLTDLSSRM
FSTKPTLIAGVGVGLIAPYEALIDSLPGTQTARLAV"
complement(7563..8960)
/gene="ml17517"
CDS
complement(7563..8960)
/gene="ml17517"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="BAB53960.1"
/db_xref="GI:14027006"
/transl_table="WQYVSTRGEAPALGFSDAVLGLADGGLYVREWPQPSASEIR
AMRGLAYPLAIRVLSPFLGGEIPAPVFERLVRAYATPRHEAVCPVLQVTGNTFVLE
LHFGPTLAFKDMAMOLLAELMDHVLAEARQHATIVGATSGDTGGAIDAFAGSRSDI
FILPHGRVSPVQOROMTTSKAENVHALAIEGNFDDCOGLKDMFNDAFRORVALSG
VNSINWARIMAQIYVYFSSALSILGAPDRPVSFTVPTGNGDIFAGVAAKMGILPIERL
VIATNDNLIARFATGRTKGVFAITSPMDIQVSSNFERLLFEASNRDAATVRRY
MDLQKSGAFTIESREINGRSEFDAGRATMDVAAATIRSTLAAASNYLLDHPHTAAMH
VAAKASGAVPMVVLGTAHPAKFPAVEAASGVSPALPAWLGMLTFESKYTVLPSPDL
KXVEDYVSRRAAR"
8930..9817
/gene="ml17518"
CDS
8930..9817
/gene="ml17518"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53961.1"
/db_xref="GI:14027007"
/transl_table="MLPRAYSHIACLSPYAAMLVSHVHKVNFCASSRIFAALIVVT
SFTRGKCSMAFRTLSGRFLAGLUTGFMALAGCOSGNGILNLGFGKQDTPAPP
QDPKILASQUNAYCPRTVTRDGTAFNTYANVKAKPKLKKTGDAADAAQAAADA
AAATPPDSARIIYQASIAIDVDRDCSRADGQAMKIAVAGKIVPGKFTPGTITMPIR
TAVMHGTDLVLSQIHQYQVQDPSVATQFVFTSDNVVVVPAPTAQDYQVYAGYDEAP
KATYTDKPKTKRKAATN"
complement(9860..10549)
/gene="ml17519"
CDS
complement(9860..10549)
/gene="ml17519"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53962.1"
/db_xref="GI:14027008"
/transl_table="MPQPDILVFDGVLVDSEIINARVEABLITLAGYEISAEETAE"
```





us-10-624-714-9.rge

Mon May 3 09:03:29 2004

Search completed: April 29, 2004, 06:01:58  
Job time : 367.465 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 1892.42 Seconds  
(without alignments)  
378.717 Million cell updates/sec

Title: US-10-624-714-9

Perfect score: 24

Sequence: 1 cgcgcagagccgatccgaagc 24

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_hrc.\*

9: gb\_estl.\*

10: gb\_est2.\*

11: gb\_hrc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gssl.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.4	89.2	538	14	CD902576
2	21.4	89.2	558	14	CD917962
C 3	21.4	89.2	647	13	BU999541
C 4	21.4	89.2	714	14	CD902577

RESULT 1  
CD902576  
LOCUS  
DEFINITION G356.107G19F010918 G356 Triticum aestivum cDNA clone G356107G19,  
mRNA sequence.  
CD902576  
VERSION CD902576.1 GI:32676904  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticaceae; Triticum.  
AUTHORS 1 (bases 1 to 538)  
TITLE Genoplatte, a major partnership french program in plant genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Genoplatte  
Genoplatte  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)  
and <http://genoplatte-info.infobiogen.fr>.  
Location/Qualifiers  
1. .538

ALIGNMENTS

C	5	20.8	86.7	222	12	BI417178
	6	20.8	86.7	264	14	CF646281
	7	20.8	86.7	382	14	CF650853
	8	20.8	86.7	447	14	CF647980
C	9	20.8	86.7	456	29	CG136115
	10	20.8	86.7	457	14	CF649202
	11	20.8	86.7	461	14	CF649860
	12	20.8	86.7	510	14	CF648527
C	13	20.8	86.7	548	12	BI398523
	14	20.8	86.7	548	12	CF244638
C	15	20.8	86.7	554	12	BI398522
	16	20.8	86.7	561	12	BI894200
C	17	20.8	86.7	570	14	CA165498
	18	20.8	86.7	584	12	BI478850
C	19	20.8	86.7	592	13	CA131069
	20	20.8	86.7	605	12	BI894199
C	21	20.8	86.7	606	14	CA192810
	22	20.8	86.7	622	12	BI478849
C	23	20.8	86.7	629	13	CA100371
	24	20.8	86.7	639	14	CF244508
	25	20.8	86.7	640	14	CA232578
	26	20.8	86.7	676	13	CA108766
	27	20.8	86.7	677	13	CA072226
	28	20.8	86.7	681	14	CA190842
	29	20.8	86.7	697	13	CA163299
C	30	20.8	86.7	703	14	CF876371
C	31	20.8	86.7	725	14	CF870291
	32	20.8	86.7	727	14	CF244232
	33	20.8	86.7	731	28	BZ707826
	34	20.8	86.7	739	29	CG049850
	35	20.8	86.7	742	14	CF244173
C	36	20.8	86.7	776	14	CB905051
C	37	20.8	86.7	795	14	CB900464
	38	20.8	86.7	796	14	CA183217
	39	20.8	86.7	823	14	CF243638
C	40	20.8	86.7	839	28	CC350392
	41	20.8	86.7	841	28	CC383940
	42	20.8	86.7	889	14	CA157469
C	43	20.8	86.7	964	14	CF884557
	44	20.8	86.7	1065	14	CA209236
	45	19.8	82.5	326	9	AV620758

CD902576 538 bp mRNA linear EST 14-JUL-2003  
G356.107G19F010918 G356 Triticum aestivum cDNA clone G356107G19,  
mRNA sequence.

CD902576.1 GI:32676904

EST.

Triticum aestivum (bread wheat)

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticaceae; Triticum.

1 (bases 1 to 538)

Genoplatte.

Genoplatte, a major partnership french program in plant genomics

Unpublished (2003)

Contact: Genoplatte

Genoplatte

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french

plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)

and <http://genoplatte-info.infobiogen.fr>.

Location/Qualifiers

1. .538

FEATURES  
source

```

/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G356107G19"
/tissue_type="grain (356 degrees per day after
pollination)"
/clone_lib="G356"

ORIGIN
Query Match      89.2%; Score 21.4; DB 14; Length 538;
Best Local Similarity 95.7%; Pred. No. 6.4e-03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAG 23
    |||
Db 421 CGCGACGAGCGCGATGCCGAAG 443

RESULT 2
CD917962      538 bp mRNA linear EST 14-JUL-2003
DEFINITION    G608.107H09F010906 G608 Triticum aestivum cDNA clone G608107H09,
              mRNA sequence.
ACCESSION     CD917962
VERSION       CD917962.1 GI:32692286
KEYWORDS      EST.
SOURCE        Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
REFERENCE     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Poideae; Triticeae; Triticum.
              1 (bases 1 to 538)
AUTHORS       Genoplante.
TITLE         Genoplante, a major partnership french program in plant genomics
JOURNAL       Unpublished (2003)
COMMENT       Contact: Genoplante
              93, rue Henri Rochefort 91025 EVRY CEDEX France
              Tel: 33 1 69 47 54 00
              Fax: 33 1 69 47 54 10
              This sequence has been generated in the framework of the french
              plant genomics programme 'genoplante' (http://www.genoplante.com
              and http://genoplante-info.infobiogen.fr).

FEATURES
source
1..558
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G608107H09"
/tissue_type="grain (608 degrees per day after
pollination)"
/clone_lib="G608"

ORIGIN
Query Match      89.2%; Score 21.4; DB 14; Length 558;
Best Local Similarity 95.7%; Pred. No. 6.4e-03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAG 23
    |||
Db 178 CGCGACGAGCGCGATGCCGAAG 200

RESULT 3
BU999541/c     647 bp mRNA linear EST 23-OCT-2002
LOCUS          H14N03r HI Hordeum vulgare subsp. vulgare cDNA clone H14N03
DEFINITION     5-PRIME, mRNA sequence.
ACCESSION     BU999541
VERSION       BU999541.1 GI:24276524
KEYWORDS      EST.

```

```

SOURCE
ORGANISM      Hordeum vulgare subsp. vulgare
              Hordeum vulgare subsp. vulgare
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Poideae; Triticeae; Hordeum.
              1 (bases 1 to 647)
AUTHORS       Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
TITLE         EST sequencing and analysis in barley (2002)
JOURNAL        Unpublished (2002)
COMMENT       Contact: Stein Nils
              Molecular Markers Group, Department Genbank
              Institute of Plant Genetics and Crop Plant Research (IPK)
              Corrensstr. 3, 06466, Gatersleben, Germany
              Tel: 039482-5522
              Fax: 039482-5595
              Email: stein@ipk-gatersleben.de
              Insert length: 647 Std Error: 0.00
              Plate: 14 row: N column: 3
              Seq primer: M13rev.
              Location/Qualifiers
              1..647
              /organism="Hordeum vulgare subsp. vulgare"
              /mol_type="mRNA"
              /cultivar="barke"
              /sub_species="vulgare"
              /db_xref="GABI:253405"
              /db_xref="taxon:112509"
              /clone="H14N03"
              /tissue_type="female inflorescences"
              /dev_stage="female inflorescences (approx. 3 mm in size)"
              /lab_host="XL10-Gold"
              /clone_lib="HI"
              /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
              cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning
              artefact caused by the kit, in most cases the EcoRI site
              is NOT present, as well as the EcoRI adapter used for
              cloning. To excise the insert, restriction sites upstream
              EcoRI should be used (e.g. BamHI, Sali, PstI). NOTE: Also
              due to the cloning system used Blue/white selection for
              recombinants is not 100% reliable."

ORIGIN
Query Match      89.2%; Score 21.4; DB 13; Length 647;
Best Local Similarity 95.7%; Pred. No. 6.5e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCGACGAGCGCGATGCCGAAG 23
    |||
Db 192 CGCGACGAGCGCGATGCCGAAG 170

RESULT 4
CD902577/c     714 bp mRNA linear EST 14-JUL-2003
LOCUS          G356.107G19R011024 G356 Triticum aestivum cDNA clone G356107G19,
DEFINITION     mRNA sequence.
ACCESSION     CD902577
VERSION       CD902577.1 GI:32676905
KEYWORDS      EST.
SOURCE        Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Poideae; Triticeae; Triticum.
              1 (bases 1 to 714)
AUTHORS       Genoplante.
TITLE         Genoplante, a major partnership french program in plant genomics
JOURNAL       Unpublished (2003)
COMMENT       Contact: Genoplante
              Genoplante
              93, rue Henri Rochefort 91025 EVRY CEDEX France
              Tel: 33 1 69 47 54 00
              Fax: 33 1 69 47 54 10

```

This sequence has been generated in the framework of the french plant genomics programme 'Genoplante' (<http://www.genoplante.com>) and <http://genoplante-info.infobiogen.fr/>.

## FEATURES

source

```
1..714
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="G356107G19"
/tissue_type="grain (356 degrees per day after
pollination)"
/clone_lib="G356"
```

## ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 714;  
Best Local Similarity 95.7%; Pred. No. 6.6e-03;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCGACGAGCGCGATGCCGAG 23  
Db 381 CGCGACGAGCGCGATGCCGAG 359

## RESULT 5

Bi417178/c

LOCUS Bi417178 222 bp mRNA linear EST 28-JAN-2002  
DEFINITION B52002D01.x9 952 - B5S tissue from Walbot Lab (reduced rRNA) Zea  
mays cDNA, mRNA sequence.

ACCESSION Bi417178

VERSION Bi417178.1 GI:15188201

KEYWORDS EST.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS

TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL

COMMENT

Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 952002 row: D column: 01.

## FEATURES

source

```
1..222
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/tissue_type="suspension culture"
/dev_stages="mixed logarithmic and stationary growth
phases"
/lab_host="DH10B"
/clone_lib="952 - BMS tissue from Walbot Lab (reduced
rRNA)"
/notes="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
library was prepared by George Rudenko using poly (A)
selected RNA and Universal Riboclone cDNA Synthesis System
(Promega). cDNA was synthesized using both random and
oligo(dT) primers in separate reactions and equipped with
EcoRI adaptors. Library was size-fractionated on agarose
gels (for insert size >400bp) and non-directionally cloned
into EcoRI-digested pUC19 vector. Blue/white selection on
carbenicillin-containing plates was used to recover
positive clones."
```

## ORIGIN

Query Match 86.7%; Score 20.8; DB 12; Length 222;  
Best Local Similarity 91.7%; Pred. No. 8.9e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCGACGAGCGCGATGCCGAGC 24  
Db 24 CGCGACGAGCGCGCGCGGAGC 1

## RESULT 6

CF646281

LOCUS

DEFINITION CF646281 264 bp mRNA linear EST 02-OCT-2003  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

ACCESSION CF646281

VERSION CF646281.1 GI:37417257

KEYWORDS EST.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS

TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL

COMMENT

Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 3530\_1\_111\_1 row: H column: 10.

## FEATURES

source

```
1..264
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="multiple"
/dev_stages="varies by tissue"
/lab_host="DH10B"
/clone_lib="3530 - Full length cDNA library created by
Invitrogen from multiple tissues"
/notes="Organ: silks, husks, ears, pollen, shoot tips,
leaf, root tips, whole seed, embryo; Vector: pCMV-Sport
6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery
Project contracted with Invitrogen to produce a
normalized, full length library in a pSport vector. This
is a Gateway compatible vector, permitting clone movement
to new vector backbones for expression in diverse host
cells using recombination rather than restriction enzymes.
Details of the vector and sequencing primers are available
at ZmDB in the EST library description tables. poly(A)+
RNA was prepared by Invitrogen, and equimolar amounts of
RNA from each of the 12 tissue samples were mixed together
for selection of mRNA with a 5' cap. After synthesis of
cDNA, a normalization step was conducted against the
mixture of RNA sources. This step effected a 20X to 80X
reduction in common transcript types. Tissues prepared: 1.
just emerging silks; 2. inner husks from ears of sample
#1; 3. 20 day aleurone; 4. immature tassels, stages from
1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm
vegetative shoot tips from 15 day old seedlings; all
leaves with an expanded or partially expanded sheath
were removed; 8. mature leaf tissue; 9. 0.5 cm long root
tips from 15 day old seedlings; 10. 10 day whole seed; 11.
12 day endosperm and embryo; 12. 17 day endosperm and
embryo. All of the sequenced clones in project 3530 will
be archived at the University of Arizona along with the
```

Unigene clones from the Maize Gene Discovery EST sequencing projects. Clones can be ordered through the ZmDB web site or directly from the University of Arizona (<http://www.genome.arizona.edu/orders/>). High density filters containing over 18,000 clones can also be ordered from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 264;  
Best Local Similarity 91.7%; Pred. No. 9.1e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGGATGCCGAGC 24  
|||||  
Db 233 CGCCGACGAGCGCGGATGCCGAGC 256  
|||||

RESULT 7  
CF650853 382 bp mRNA linear EST 02-OCT-2003  
LOCUS 3530.1.94.1 C02.x.1 3530 - Full length cDNA library created by  
DEFINITION Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

ACCESSION CF650853  
VERSION CF650853.1 GI:37426204

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 382)

## REFERENCE

## AUTHORS

## TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

## JOURNAL

## COMMENT

Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 3530.1.94.1 row: C column: 02.

## FEATURES

## source

1..382  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/tissue\_type="multiple"  
/dev\_stage="varies by tissue"  
/lab\_host="DH10B"  
/clone\_lib="3530 - Full length cDNA library created by  
Invitrogen from multiple tissues"  
/note="Organ: silks, husks, ears, pollen, shoot tips,  
leaf, root tips, whole seed, embryo; Vector: pCMV-Sport  
6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery  
Project contracted with Invitrogen to produce a  
normalized, full length library in a sport vector. This  
is a Gateway compatible vector, permitting clone movement  
to new vector backbones for expression in diverse host  
cells using recombination rather than restriction enzymes.  
Details of the vector and sequencing primers are available  
at ZmDB in the EST library description tables. poly(A)+  
mRNA was prepared by Invitrogen, and equimolar amounts of  
RNA from each of the 12 tissue samples were mixed together  
for selection of mRNA with a 5' cap. After synthesis of  
cDNA, a normalization step was conducted against the  
mixture of RNA sources. This step effected a 20X to 80X  
reduction in common transcript types. Tissues prepared: 1.  
just emerging silks; 2. inner husks from ears of sample  
#1; 3. 20 dap aleurone; 4. immature tassels, stages from  
1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm

vegetative shoot tips from 15 day old seedlings; all  
leaves with an expanded or partially expanded sheath  
were removed; 8. mature leaf tissue; 9. 0.5 cm long root  
tips from 15 day old seedlings; 10. 10 dap whole seed; 11.  
12 dap endosperm and embryo; 12. 17 dap endosperm and  
embryo. All of the sequenced clones in project 3530 will  
be archived at the University of Arizona along with the  
Unigene clones from the Maize Gene Discovery EST  
sequencing projects. Clones can be ordered through the  
ZmDB web site or directly from the University of Arizona  
(<http://www.genome.arizona.edu/orders/>). High density  
filters containing over 18,000 clones can also be ordered  
from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 382;  
Best Local Similarity 91.7%; Pred. No. 9.4e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGGATGCCGAGC 24  
|||||  
Db 349 CGCCGACGAGCGCGGATGCCGAGC 372  
|||||

## RESULT 8

## CF647980

## LOCUS

## DEFINITION

3530.1.50.1 E09.x.1 3530 - Full length cDNA library created by

## ACCESSION

## CF647980

## VERSION

## CF647980.1

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## Unpublished (1999)

## Contact: Walbot V

## Department of Biological Sciences

## Stanford University

## 855 California Ave, Palo Alto, CA 94304, USA

## Tel: 650 723 2227

## Fax: 650 725 8221

## Email: walbot@stanford.edu

## Plate: 3530.1.50.1 row: E column: 09.

## Location/Qualifiers

## 1..447

## /organism="Zea mays"

## /mol\_type="mRNA"

## /cultivar="B73"

## /db\_xref="taxon:4577"

## /tissue\_type="multiple"

## /dev\_stage="varies by tissue"

## /lab\_host="DH10B"

## /clone\_lib="3530 - Full length cDNA library created by

## Invitrogen from multiple tissues"

## /note="Organ: silks, husks, ears, pollen, shoot tips,

## leaf, root tips, whole seed, embryo; Vector: pCMV-Sport

## 6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery

## Project contracted with Invitrogen to produce a

## normalized, full length library in a sport vector. This

## is a Gateway compatible vector, permitting clone movement

## to new vector backbones for expression in diverse host

## cells using recombination rather than restriction enzymes.

## Details of the vector and sequencing primers are available

## at ZmDB in the EST library description tables. poly(A)+

## mRNA was prepared by Invitrogen, and equimolar amounts of

## RNA from each of the 12 tissue samples were mixed together

## for selection of mRNA with a 5' cap. After synthesis of

## cDNA, a normalization step was conducted against the

## mixture of RNA sources. This step effected a 20X to 80X

## reduction in common transcript types. Tissues prepared: 1.

## just emerging silks; 2. inner husks from ears of sample

## #1; 3. 20 dap aleurone; 4. immature tassels, stages from

## 1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm

for selection of mRNA with a 5' cap. After synthesis of cDNA, a normalization step was conducted against the mixture of RNA sources. This step effected a 20X to 80X reduction in common transcript types. Tissues prepared: 1. just emerging silks; 2. inner husks from ears of sample #1; 3. 20 dap aleurone; 4. immature tassels, stages from 1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm vegetative shoot tips from 15 day old seedlings; all leaves with an expanded or partially expanded sheath were removed; 8. mature leaf tissue; 9. 0.5 cm long root tips from 15 day old seedlings; 10. 10 dap whole seed; 11. 12 dap endosperm and embryo; 12. 17 dap endosperm and embryo. All of the sequenced clones in project 3530 will be archived at the University of Arizona along with the Unigene clones from the Maize Gene Discovery EST ZmDB web site or directly from the University of Arizona (<http://www.genome.arizona.edu/orders/>). High density filters containing over 18,000 clones can also be ordered from the University of Arizona.

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 447;  
 Best Local Similarity 91.7%; Pred. No. 9.5e+03;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
 |||||  
 Db 227 CGCCGACGAGCGCGCGCGCGAGC 250  
 |||||

## RESULT 9

CG136115/c  
 LOCUS CG136115  
 DEFINITION PUIDL27TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMVBTA0559E06,  
 genomic survey sequence.

ACCESSION CG136115  
 VERSION CG136115.1 GI:34025167  
 KEYWORDS GSS.

SOURCE Zea mays  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 456)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.

Maize Genomics Consortium  
 Unpublished (2003)

Other GSSs: PUIDL27TB

Contact: Cathy Whitelaw

## TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

## FEATURES

source

1..456  
 Location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMVBTA0559E06"  
 /clone\_lib="ZM\_0.6\_1.0\_KB"  
 /note="Vector: pCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

## ORIGIN

Query Match 86.7%; Score 20.8; DB 29; Length 456;  
 Best Local Similarity 91.7%; Pred. No. 9.5e+03;

## Matches

22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY

1 CGCCGACGAGCGCGATGCCGAGC 24

## Db

182 CGCCGACGAGCGCGCGCGAGC 159

## RESULT 10

CF649202  
 LOCUS CF649202  
 DEFINITION 3530.1.66\_1.B02.x.1.3530 - Full length cDNA library created by  
 Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

## ACCESSION

CF649202

## VERSION

CF649202.1 GI:37422968

## KEYWORDS

EST.

## SOURCE

Ze mays

## ORGANISM

Ze mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 457)

Walbot,V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3530.1.66.1 row: B column: 02.

Location/Qualifiers

1..457

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/tissue\_type="multiple"

/dev\_stage="varies by tissue"

/lab\_host="DH10B"

/clone\_lib="3530 - Full length cDNA library created by

Invitrogen from multiple tissues"

/note="Organ: silks, husks, ears, pollen, shoot tips

leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT

6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery

Project contracted with Invitrogen to produce a

normalized, full length library in a pSPORT vector. This

is a Gateway compatible vector, permitting clone movement

to new vector backbones for expression in diverse host

cells using recombination rather than restriction enzymes.

Details of the vector and sequencing primers are available

at ZmDB in the EST library description tables. poly(A)+

mRNA was prepared by Invitrogen, and equimolar amounts of

RNA from each of the 12 tissue samples were mixed together

for selection of mRNA with a 5' cap. After synthesis of

cDNA, a normalization step was conducted against the

mixture of RNA sources. This step effected a 20X to 80X

reduction in common transcript types. Tissues prepared: 1.

just emerging silks; 2. inner husks from ears of sample

#1; 3. 20 dap aleurone; 4. immature tassels, stages from

1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm

vegetative shoot tips from 15 day old seedlings; all

leaves with an expanded or partially expanded sheath

were removed; 8. mature leaf tissue; 9. 0.5 cm long root

tips from 15 day old seedlings; 10. 10 dap whole seed; 11.

12 dap endosperm and embryo; 12. 17 dap endosperm and

embryo. All of the sequenced clones in project 3530 will

be archived at the University of Arizona along with the

Unigene clones from the Maize Gene Discovery EST

sequencing projects. Clones can be ordered through the

ZmDB web site or directly from the University of Arizona

(http://www.genome.arizona.edu/orders/). High density filters containing over 18,000 clones can also be ordered from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 457;  
Best Local Similarity 91.7%; Pred. No. 9.5e+03;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 398 CGCCGACGAGCGCGCGCGCGAGC 421  
|||||

RESULT 11 461 bp mRNA linear EST 02-OCT-2003  
CF649860  
LOCUS 3530.1\_77.1\_D03.x.1 3530 - Full length cDNA library created by  
DEFINITION Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.  
CF649860  
ACCESSION CF649860.1 GI:37424257

VERSION  
KEYWORDS  
SOURCE

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 461)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3530.1\_77.1 row: D column: 03.

## FEATURES

## source

1. .461

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/tissue\_type="multiple"

/dev\_stage="varies by tissue"

/lab\_host="DH10B"

/clone\_lib="3530 - Full length cDNA library created by

Invitrogen from multiple tissues"

/note="Organ: silks, husks, ears, pollen, shoot tips,

leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT

6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery

Project contracted with Invitrogen to produce a

normalized, full length library in a pSport vector. This

is a Gateway compatible vector, permitting clone movement

to new vector backbones for expression in diverse host

cells using recombination rather than restriction enzymes.

Details of the vector and sequencing primers are available

at ZmDB in the EST library description tables. poly(A)+

mRNA was prepared by Invitrogen, and equimolar amounts of

RNA from each of the 12 tissue samples were mixed together

for selection of mRNA with a 5' cap. After synthesis of

cDNA, a normalization step was conducted against the

mixture of RNA sources. This step effected a 20X to 80X

reduction in common transcript types. Tissues prepared: 1.

Just emerging silks; 2. Inner husks from ears of sample

#1; 3. 20 day aleurone; 4. Immature tassels, stages from

1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm

vegetative shoot tips from 15 day old seedlings; all

leaves with an expanded or partially expanded sheath

were removed; 8. mature leaf tissue; 9. 0.5 cm long root

tips from 15 day old seedlings; 10. 10 day whole seed; 11.  
12 day endosperm and embryo; 12. 17 day endosperm and  
embryo. All of the sequenced clones in project 3530 will  
be archived at the University of Arizona along with the  
Unigene clones from the Maize Gene Discovery EST  
sequencing projects. Clones can be ordered through the  
ZmDB web site or directly from the University of Arizona  
(http://www.genome.arizona.edu/orders/). High density  
filters containing over 18,000 clones can also be ordered  
from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 461;  
Best Local Similarity 91.7%; Pred. No. 9.5e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 241 CGCCGACGAGCGCGCGCGCGAGC 264  
|||||

## RESULT 12

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

CF648527 510 bp mRNA linear EST 02-OCT-2003  
3530.1\_56.1\_C05.x.1 3530 - Full length cDNA library created by  
Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

CF648527

CF648527.1 GI:37421653

EST.

KEYWORDS

SOURCE

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 510)

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3530.1\_56.1 row: C column: 05.

Location/Qualifiers

1. .510

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/tissue\_type="multiple"

/dev\_stage="varies by tissue"

/lab\_host="DH10B"

/clone\_lib="3530 - Full length cDNA library created by

Invitrogen from multiple tissues"

/note="Organ: silks, husks, ears, pollen, shoot tips,

leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT

6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery

Project contracted with Invitrogen to produce a

normalized, full length library in a pSport vector. This

is a Gateway compatible vector, permitting clone movement

to new vector backbones for expression in diverse host

cells using recombination rather than restriction enzymes.

Details of the vector and sequencing primers are available

at ZmDB in the EST library description tables. poly(A)+

mRNA was prepared by Invitrogen, and equimolar amounts of

RNA from each of the 12 tissue samples were mixed together

for selection of mRNA with a 5' cap. After synthesis of

cDNA, a normalization step was conducted against the

mixture of RNA sources. This step effected a 20X to 80X



reduction in common transcript types. Tissues prepared: 1. just emerging silks; 2. inner husks from ears of sample #1; 3. 20 dap aleurone; 4. immature tassels, stages from 1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm vegetative shoot tips from 15 day old seedlings; all leaves with an expanded or partially expanded sheath were removed; 8. mature leaf tissue; 9. 0.5 cm long root tips from 15 day old seedlings; 10. 10 dap whole seed; 11. 12 dap endosperm and embryo; 12. 17 dap endosperm and embryo. All of the sequenced clones in project 3530 will be archived at the University of Arizona along with the Unigene clones from the Maize Gene Discovery EST sequencing projects. Clones can be ordered through the ZmDB web site or directly from the University of Arizona (<http://www.genome.arizona.edu/orders/>). High density filters containing over 18,000 clones can also be ordered from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 510;  
Best Local Similarity 91.7%; Pred. No. 9.6e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24

Db 233 CGCCGACGAGCGCGCGCGCGAGC 256

## RESULT 13

BI398523/c 548 bp mRNA linear EST 28-JAN-2002  
LOCUS 952002D01.x8 952 - BMS tissue from Walbot Lab (reduced rRNA) Zea  
DEFINITION mays cDNA, mRNA sequence.

ACCESSION BI398523

VERSION BI398523.1 GI:15177584

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 548)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 952002 row: D column: 01.

Location/Qualifiers

## FEATURES

source

1..548  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="BMS (Black Mexican Sweet)"  
/db\_xref="taxon:4577"  
/tissue\_type="suspension culture"  
/dev\_stage="mixed logarithmic and stationary growth phases"  
/lab\_host="DH10B"  
/clone\_lib="952 - BMS tissue from Walbot Lab (reduced rRNA)"  
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned

into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 12; Length 548;  
Best Local Similarity 91.7%; Pred. No. 9.7e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24

Db 280 CGCCGACGAGCGCGCGCGAGC 257

## RESULT 14

CF244638 548 bp mRNA linear EST 06-AUG-2003  
LOCUS 3530.1.3.1.H11.Y.1.3530 - Full length cDNA library created by  
DEFINITION Invitrogen from multiple tissues Zea mays cDNA, mRNA sequence.

ACCESSION CF244638

VERSION CF244638.1 GI:33467589

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 548)

AUTHORS Walbot,V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

JOURNAL University

COMMENT Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3530.1.3.1 row: H column: 11.

Location/Qualifiers

## FEATURES

source

1..548  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/tissue\_type="multiple"  
/dev\_stage="varies by tissue"  
/lab\_host="DH10B"  
/clone\_lib="3530 - Full length cDNA library created by  
Invitrogen from multiple tissues"  
/note="Organ: silks, husks, ears, pollen, shoot tips,  
leaf, root tips, whole seed, embryo; Vector: pCMV-SPORT  
6.1; Site 1: EcoRV; Site 2: NotI; Maize Gene Discovery  
Project contracted with Invitrogen to produce a  
normalized, full length library in a pSport vector. This  
is a Gateway compatible vector, permitting clone movement  
to new vector backbones for expression in diverse host  
cells using recombination rather than restriction enzymes.  
Details of the vector and sequencing primers are available  
at ZmDB in the EST library description tables. poly(A)+  
mRNA was prepared by Invitrogen, and equimolar amounts of  
RNA from each of the 12 tissue samples were mixed together  
for selection of mRNA with a 5' cap. After synthesis of  
cDNA, a normalization step was conducted against the  
mixture of RNA sources. This step effected a 20X to 80X  
reduction in common transcript types. Tissues prepared: 1.  
just emerging silks; 2. inner husks from ears of sample  
#1; 3. 20 dap aleurone; 4. immature tassels, stages from  
1-2 mm to 1-2 cm; 5. 2 mm to 2 cm ears; 6. pollen; 7. 1 cm  
vegetative shoot tips from 15 day old seedlings; all  
leaves with an expanded or partially expanded sheath  
were removed; 8. mature leaf tissue; 9. 0.5 cm long root

tips from 15 day old seedlings; 10, 10 dap whole seed; 11, 12 dap endosperm and embryo; 12, 17 dap endosperm and embryo. All of the sequenced clones in project 3530 will be archived at the University of Arizona along with the Unigene clones from the Maize Gene Discovery EST sequencing projects. Clones can be ordered through the ZmDB web site or directly from the University of Arizona (<http://www.genome.arizona.edu/orders/>). High density filters containing over 18,000 clones can also be ordered from the University of Arizona."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 14; Length 548;  
Best Local Similarity 91.7%; Pred. No. 9.7e+03;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 218 CGCCGACGAGCGCGCGCGGAGC 241  
|||||

## RESULT 15

BI398522/c  
LOCUS BI398522 554 bp mRNA linear EST 28-JAN-2002  
DEFINITION 952002D01.x7 952 - BMS tissue from Walbot Lab (reduced rRNA) Zea.  
mays cDNA, mRNA sequence.

ACCESSION BI398522  
VERSION BI398522.1 GI:15177583  
KEYWORDS EST.

## SOURCE

Zea mays

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

Walbot.V.  
Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 952002 row: D column: 01.

## FEATURES

## source

1..554  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="BMS (Black Mexican Sweet)"  
/db\_xref="taxon:4577"  
/tissue\_type="suspension culture"  
/dev\_stage="mixed logarithmic and stationary growth  
phases"  
/lab\_host="DH10B"  
/clone\_lib="952 - BMS tissue from Walbot Lab (reduced  
rRNA)"  
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The  
library was prepared by George Rudenko using poly (A)  
selected RNA and Universal Riboclone cDNA Synthesis System  
(Promega). cDNA was synthesized using both random and  
oligo(dT) primers in separate reactions and equipped with  
EcoRI adaptors. Library was size-fractionated on agarose  
gels (for insert size >400bp) and non-directionally cloned  
into EcoRI-digested pUC19 vector. Blue/white selection on  
carbenicillin-containing plates was used to recover  
positive clones."

## ORIGIN

Query Match 86.7%; Score 20.8; DB 12; Length 554;  
Best Local Similarity 91.7%; Pred. No. 9.7e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 280 CGCCGACGAGCGCGCGCGGAGC 257  
|||||

Search completed: April 29, 2004, 11:37:13  
Job time : 1896.42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 42.3529 Seconds  
(without alignments)  
314.472 Million cell updates/sec

Title: US-10-624-714-9

Perfect score: 24

Sequence: 1 cgccgacgagcgatgccgaagc 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	4403765	3	US-09-103-840A-2
2	24	100.0	4411529	3	US-09-103-840A-1
3	19.8	82.5	1194	4	US-09-489-039A-3387
4	18.2	80.0	405	4	US-09-252-991A-4225
5	19.2	80.0	807	4	US-09-252-991A-164
6	19.2	80.0	843	4	US-09-252-991A-4551
7	19.2	80.0	891	4	US-09-252-991A-176
8	19.2	80.0	948	4	US-09-252-991A-172
9	19.2	80.0	1128	4	US-09-252-991A-4450
10	19.2	80.0	1242	4	US-09-252-991A-4304
11	19.2	80.0	1302	4	US-09-252-991A-4157
12	19.2	80.0	1611	4	US-09-252-991A-183
13	18.8	78.3	68750	3	US-09-335-409-1
14	18.8	78.3	68750	4	US-09-568-102-1
15	18.8	78.3	68750	4	US-09-567-969-1
16	18.8	78.3	68750	4	US-09-568-480-1
17	18.8	78.3	68750	4	US-09-568-486-1
18	18.8	78.3	68750	4	US-09-568-472-1
19	18.8	78.3	68750	4	US-09-567-899-1
20	18.8	78.3	71989	4	US-09-443-501A-2
21	18.4	76.7	236	4	US-09-480-251-5
22	18.4	76.7	488	4	US-09-480-251-1
23	18.2	75.8	606	4	US-09-252-991A-11629
24	18.2	75.8	945	4	US-09-252-991A-11511
25	18.2	75.8	1128	4	US-09-252-991A-3879
26	18.2	75.8	1731	4	US-09-252-991A-4008
27	18.2	75.8	2472	4	US-09-252-991A-3902

28 18.2 75.8 2562 4 US-09-252-991A-4045 Sequence 4045, Ap  
c 29 18.2 75.8 4326 4 US-09-252-991A-11572 Sequence 11572, A  
30 18.2 75.8 4473 4 US-09-252-991A-11871 Sequence 11871, A  
c 31 18.2 75.8 4403765 3 US-09-103-840A-2 Sequence 2, Appli  
c 32 18.2 75.8 4411529 3 US-09-103-840A-1 Sequence 1, Appli  
33 17.8 74.2 638 3 US-08-998-416-981 Sequence 981, App  
34 17.8 74.2 1187 1 US-08-440-856A-2 Sequence 2, Appli  
35 17.8 74.2 1287 4 US-09-489-039A-6998 Sequence 6998, Ap  
c 36 17.8 74.2 1737 4 US-09-489-039A-6926 Sequence 6926, Ap  
c 37 17.8 74.2 18318 1 US-08-414-926A-6 Sequence 6, Appli  
c 38 17.8 74.2 18318 2 US-08-926-922-6 Sequence 6, Appli  
c 39 17.8 74.2 18318 3 US-09-253-682-6 Sequence 6, Appli  
c 40 17.8 74.2 18318 3 US-09-527-657-6 Sequence 6, Appli  
c 41 17.8 74.2 18318 4 US-09-892-100-6 Sequence 6, Appli  
c 42 17.8 74.2 28804 2 US-08-592-874-1 Sequence 1, Appli  
c 43 17.8 74.2 28804 3 US-09-096-942-2 Sequence 2, Appli  
c 44 17.8 74.2 28804 3 US-09-096-867-2 Sequence 2, Appli  
45 17.6 73.3 216 4 US-09-252-991A-3476 Sequence 3476, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Query Match Best Local Similarity 100.0%; Score 24; DB 3; Length 4403765;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAGC 24

Db 1112875 CGCCGACGAGCGCGATGCCGAGC 1112898

##### RESULT 2

US-09-103-840A-1

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2



;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 176  
;; LENGTH: 891  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-176

Query Match 80.0%; Score 19.2; DB 4; Length 891;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 652 CGCCGCGAGCGCGAGCGGAGC 629

## RESULT 8

US-09-252-991A-172/c  
;; Sequence 172, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 172  
;; LENGTH: 948  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-172

Query Match 80.0%; Score 19.2; DB 4; Length 948;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 551 CGCCGCGAGCGCGAGCGGAGC 528

## RESULT 9

US-09-252-991A-4450  
;; Sequence 4450, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 4450  
;; LENGTH: 1128

;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4450

Query Match 80.0%; Score 19.2; DB 4; Length 1128;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 183 CGCCGCGAGCGAGCGATGCCGATGC 206

## RESULT 10

US-09-252-991A-4304/c  
;; Sequence 4304, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 4304  
;; LENGTH: 1242  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4304

Query Match 80.0%; Score 19.2; DB 4; Length 1242;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 971 CGCCGCGAGCGAGCGATGCCGATGC 948

## RESULT 11

US-09-252-991A-4157/c  
;; Sequence 4157, Application US/09252991A  
;; Patent No. 6551795  
;; GENERAL INFORMATION:  
;; APPLICANT: Marc J. Rubenfield et al.  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 4157  
;; LENGTH: 1302  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4157

Query Match 80.0%; Score 19.2; DB 4; Length 1302;  
Best Local Similarity 87.5%; Pred. No. 42;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCCGACGAGCGCGATGCCGAGC 24  
|||||  
Db 1000 CGCCGCGAGCGAGCGATGCCGATGC 977



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 448 Seconds  
(Without alignments)  
242.328 Million cell updates/sec

Title: US-10-624-714-9

Perfect score: 24

Sequence: 1 CGCGAGAGGCGGCGGAGC 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.8	86.7	741	16	US-10-369-493-42402
2	19.8	82.5	1137	16	Sequence 42402, A
3	19.8	82.5	2268	16	Sequence 44493, A
4	19.4	80.8	63158	15	Sequence 1199, App
5	19.2	80.0	506	16	Sequence 1, Appl
6	19.2	80.0	543	9	US-10-292-198-1
7	19.2	80.0	554	13	US-10-260-238-199
8	19.2	80.0	554	13	Sequence 786, App
9	19.2	80.0	577	16	Sequence 771, App
10	19.2	80.0	699	16	US-10-425-114-771
11	19.2	80.0	887	16	Sequence 390, App
12	19.2	80.0	892	13	Sequence 3280, A
13	19.2	80.0	1191	16	US-10-369-493-32280
14	19.2	80.0	1560	15	Sequence 574, App
15	19.2	80.0	1575	15	Sequence 4364, App
16	19.2	80.0	1575	15	Sequence 40727, A
17	19.2	80.0	1575	15	Sequence 44493, Application US/10369493
18	19.2	80.0	1575	15	Sequence 247, App
19	19.2	80.0	1575	15	Sequence 3466, App

15	19.2	80.0	1680	13	US-10-425-114-1524	Sequence 1524, App
16	19.2	80.0	2021	13	US-10-425-114-15014	Sequence 14014, A
17	19.2	80.0	9025608	15	US-10-156-761-1	Sequence 1, Appl
18	18.8	78.3	629	9	US-09-772-656-1	Sequence 1, Appl
19	18.8	78.3	629	13	US-10-396-199A-1	Sequence 5, Appl
20	18.8	78.3	987	9	US-09-772-656-5	Sequence 9, Appl
21	18.8	78.3	987	9	US-09-772-656-9	Sequence 5, Appl
22	18.8	78.3	987	13	US-10-396-199A-5	Sequence 5, Appl
23	18.8	78.3	987	13	US-10-396-199A-9	Sequence 9, Appl
24	18.8	78.3	1068	13	US-10-425-114-22507	Sequence 22507, A
25	18.8	78.3	1089	13	US-10-424-599-115877	Sequence 115877, A
26	18.8	78.3	1104	13	US-10-425-114-34604	Sequence 34604, A
27	18.8	78.3	1645	16	US-10-310-154-197	Sequence 197, App
28	18.8	78.3	68750	14	US-10-014-717-1	Sequence 1, Appl
29	18.8	78.3	71989	13	US-09-727-889-2	Sequence 2, Appl
30	18.4	76.7	236	15	US-10-180-247-5	Sequence 5, Appl
31	18.4	76.7	488	15	US-10-180-247-1	Sequence 1, Appl
32	18.4	76.7	894	13	US-10-282-122A-25434	Sequence 25434, A
33	18.4	76.7	907	13	US-10-425-114-30146	Sequence 30146, A
34	18.4	76.7	2000	16	US-10-260-238-1874	Sequence 1874, App
35	18.2	75.8	240	9	US-09-923-876-2471	Sequence 2471, App
36	18.2	75.8	240	11	US-09-923-876-2471	Sequence 2471, App
37	18.2	75.8	253	9	US-09-923-876-1978	Sequence 1978, App
38	18.2	75.8	253	11	US-09-923-876-1978	Sequence 1978, App
39	18.2	75.8	261	9	US-09-294-093B-2396	Sequence 2396, App
40	18.2	75.8	272	9	US-09-294-093B-4035	Sequence 4035, App
41	18.2	75.8	451	13	US-10-425-114-36027	Sequence 36027, A
42	18.2	75.8	536	13	US-10-425-114-4930	Sequence 4930, App
43	18.2	75.8	546	16	US-10-260-238-194	Sequence 194, App
44	18.2	75.8	600	13	US-10-425-114-23968	Sequence 23968, A
45	18.2	75.8	651	15	US-10-156-761-2333	Sequence 2333, App

#### ALIGNMENTS

#### RESULT 1

US-10-369-493-42402  
; Sequence 42402, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 42402  
; LENGTH: 741  
; TYPE: DNA  
; ORGANISM: Halobacterium sp. NRC-1  
US-10-369-493-42402

Query Match 86.7%; Score 20.8; DB 16; Length 741;  
Best Local Similarity 91.7%; Pred. No. 9.6;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 CGCGAGAGGCGGCGGAGC 24

Db 135 CGCGAGAGGCGGCGGAGC 158

#### RESULT 2

US-10-369-493-44493  
; Sequence 44493, Application US/10369493  
; Publication No. US20030233675A1

```

; NAME/KEY: N_region
; LOCATION: (408)..(408)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (429)..(429)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (433)..(433)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (435)..(435)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (627)..(627)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1324)..(1324)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1465)..(1465)
; OTHER INFORMATION: n = any nucleotide
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1199

Query Match      82.5%; Score 19.8; DB 16; Length 2268;
Best Local Similarity 91.3%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGCCGACGAGGCCGATGCCGAG 23
DB      1059 CGACGACGAGGACGATGCCGAG 1081

RESULT 4
US-10-292-198-1/c
; Sequence 1, Application US/10292198
; Publication No. US20030157654A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben.
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: 054030-0007
; CURRENT APPLICATION NUMBER: US/10/292,198
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 10/159,257
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: US 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 63158
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
US-10-292-198-1

Query Match      80.8%; Score 19.4; DB 15; Length 63158;
Best Local Similarity 95.2%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CGCCGACGAGGCCGATGCCGA 21
DB      2249 CGCCGACGAGGCCGAGCCGA 2229

```



```
RESULT 5
US-10-260-238-199/c
; Sequence 199, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 199
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Oryza sativa
; NAME/KEY: N region
; LOCATION: (369)..(369)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-199

Query Match      80.0%; Score 19.2; DB 16; Length 506;
Best Local Similarity 87.5%; Pred. No. 49;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CGCCGACGAGCGCGATGCCGAAGC 24
      |||||
DB      359 CGCCGACGAGCGCGATGCCGAAGC 336

RESULT 6
US-09-887-576-786/c
; Sequence 786, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 786
; LENGTH: 543
; TYPE: DNA
```

```
; ORGANISM: Oryza sativa
US-09-887-576-786

Query Match      80.0%; Score 19.2; DB 9; Length 543;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CGCCGACGAGCGCGATGCCGAAGC 24
      |||||
DB      352 CGCCGACGAGCGCGATGCCGAAGC 329

RESULT 7
US-10-425-114-771/c
; Sequence 771, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 771
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700092088_FLI
US-10-425-114-771

Query Match      80.0%; Score 19.2; DB 13; Length 554;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CGCCGACGAGCGCGATGCCGAAGC 24
      |||||
DB      80  CGCCGACGAGCGCGATGCCGAAGC 57

RESULT 8
US-10-260-238-390
; Sequence 390, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
```

```
; SEQ ID NO 390
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (5)..(6)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N region
; LOCATION: (17)..(17)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-390

Query Match      80.0%; Score 19.2; DB 16; Length 577;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
Db 484 CGCCGACGAGCGCGATGCCGAAGC 507

RESULT 9
US-10-369-493-32280
; Sequence 32280, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32280
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Ralstonia metallidurans
US-10-369-493-32280

Query Match      80.0%; Score 19.2; DB 16; Length 699;
Best Local Similarity 87.5%; Pred. No. 47;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
Db 156 CGCCGACGAGCGCGATGCCGAAGC 179

RESULT 10
US-10-259-194A-574
; Sequence 574, Application US/10259194A
; Publication No. US2004010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Chassemian, Majid
; APPLICANT: Briggs, Steven F.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
```

```
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 574
; LENGTH: 887
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (662)..(682)
; OTHER INFORMATION: n = any nucleotide
US-10-259-194A-574

Query Match      80.0%; Score 19.2; DB 16; Length 887;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
Db 590 CGCCGACGAGCGCGATGCCGAAGC 613

RESULT 11
US-10-425-114-4364
; Sequence 4364, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 4364
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700355402_FLI
US-10-425-114-4364

Query Match      80.0%; Score 19.2; DB 13; Length 892;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
Db 555 CGCCGACGAGCGCGATGCCGAAGC 578

RESULT 12
US-10-369-493-40727
; Sequence 40727, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```

```
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40727
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40727

Query Match      80.0%; Score 19.2; DB 16; Length 1191;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
DB 659 CGCCGCGCTGCCGATGCCGAAGC 682

RESULT 13
US-10-259-165-247/c
; Sequence 247, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 247
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-247

Query Match      80.0%; Score 19.2; DB 15; Length 1560;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
DB 628 CGCCGCGAGCGCGAGGCCGATGC 605

RESULT 14
US-10-156-761-3466/c
; Sequence 3466, Application US/10156761
; Publication No. US20030119018A1
```

```
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HAYTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3466
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1575)
US-10-156-761-3466

Query Match      80.0%; Score 19.2; DB 15; Length 1575;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
DB 1240 CGCCGACGATCGGATGCCGAGGC 1217

RESULT 15
US-10-425-114-1524
; Sequence 1524, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingsong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1524
; LENGTH: 1680
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700156359_FLI
US-10-425-114-1524

Query Match      80.0%; Score 19.2; DB 13; Length 1680;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCCGACGAGCGCGATGCCGAAGC 24
    ||||| ||||| ||||| ||||| |||||
DB 881 CGCCGCGAGCGCGAGGCCGAGGC 904
```

Search completed: April 29, 2004, 20:45:02  
Job time : 453 secs

This Page Blank (uspto)

CC genome. The DNA sequences provide further opportunities to find antigenic  
CC or immunogenic proteins which are more effective in vaccines than the  
CC outer membrane proteins currently used  
XX  
SQ Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 3; Length 349980;  
Best Local Similarity 87.5%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGCGGGCGGAGCCGACACGGA 24  
DB 189984 CGCGGGCGGAGCCGACACGGA 189007

RESULT 9  
AAH02144/c  
ID AAH02144 standard; DNA; 889 BP.  
XX  
AC AAH02144;  
XX  
DT 24-JUL-2001 (first entry)  
XX  
DE Pseudomonas putida nucleotide sequence SEQ ID NO:2137.  
XX  
KW Species specific; genus specific; family specific; probe; detection;  
KW identification; algal; archaeal; bacterial; fungal; parasitological;  
KW microorganism; diagnosis; translation elongation factor Tu; toxin;  
KW translation elongation factor G; RecA recombinase; resistance;  
KW catalytic subunit of proton-translocating ATPase; antimicrobial; vaccine;  
KW primer; ds.  
XX  
OS Pseudomonas putida.  
XX  
PN WO200123604-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-CA001150.  
XX  
PR 28-SEP-1999; 99CA-02283458.  
PR 19-MAY-2000; 2000CA-02307010.  
XX  
PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.  
XX  
PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;  
PI Picard FU, Roy PH;  
XX  
DR WPI; 2001-245006/25.

XX Nucleic acid sequences are used to generate universal probes and primers  
XX which can be used to identify and detect the presence of algal, archaeal,  
XX bacterial, fungal and parasitological species in a test sample.  
XX  
PS Disclosure; Page 1506; 1580pp; English.  
XX  
CC The present invention describes a method for generating a repository of  
CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes  
CC and/or primers are derived. The method comprises amplifying the nucleic  
CC acids of determined algal, archaeal, bacterial, fungal and parasitological  
CC species with a combination of defined primer pairs. The method can be  
CC used for producing probes and/or primers for detecting one or more  
CC related microorganisms e.g. algae, archaea, bacteria, fungi and  
CC parasites, for universal detection and for specific and ubiquitous  
CC detection and identification of an algal, archaeal, bacterial, fungal and  
CC parasitological species, genus, family and group. A nucleic acid (I) obtained  
CC using the method of the invention can be used for the universal detection  
CC of any bacterium, fungus or parasite in a sample and for the detection of  
CC at least one antimicrobial agent resistance gene or at least one toxin  
CC gene. hexA nucleic acids are used for the specific and ubiquitous  
CC detection and for identification of Streptococcus pneumoniae. (I) can be  
CC used to design a therapeutic agent which is effective against  
CC microorganisms. Microbial species or genus or family or phylum or group

CC which can be detected include Abiotrophia adiacens, Bordetella sp.,  
CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,  
CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp., Neisseria  
CC gonorrhoeae and Staphylococcus sp. Using DNA based tests provides faster  
CC results than substrate specificity tests as results can be determined in  
CC an hour and improved accuracy is also achieved. AAH00010 to AAH002304  
CC represent nucleotide sequences and primers/probes which are given in the  
CC exemplification of the present invention  
XX  
SQ Sequence 889 BP; 183 A; 260 C; 256 G; 189 T; 0 U; 1 Other;

Query Match 75.2%; Score 18.8; DB 4; Length 889;  
Best Local Similarity 90.9%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 CGCGGGCGGAGCCGACACGAGAT 25  
DB 55 CGCGGGCGGAGCCGACACGAGAT 34

RESULT 10  
ACA43747/c  
ID ACA43747 standard; DNA; 918 BP.  
XX  
AC ACA43747;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #25404.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX  
OS Pseudomonas putida.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Travick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR P-PSDB; ABU39877.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 31617; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 918 BP; 193 A; 261 C; 266 G; 198 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 7; Length 918;  
 Best Local Similarity 90.9%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 CGGCGGCGAGCCAAACCCAGGAT 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 55 CGGCGGCGGAGCAAAACCCAGGAT 34

RESULT 11  
 ACA43762/c  
 ID ACA43762 standard; DNA; 1191 BP.  
 XX  
 AC ACA43762;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #25419.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS *Pseudomonas putida*.  
 XX  
 PN W0200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX

XX 21-MAR-2002; 2002WO-US0009107.  
 XX  
 XX 21-MAR-2001; 2001US-00815242.  
 XX  
 PR 06-SEP-2001; 2001US-00948993.  
 PR  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR  
 PR 08-FEB-2002; 2002US-00072851.  
 PR  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI; 2003-029926/02.  
 DR  
 DR P-PSDB; ABU39892.  
 XX  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 XX for homologous nucleic acids required for cellular proliferation to  
 XX isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 14; SEQ ID NO 31632; 1766pp; English.  
 PS  
 XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1191 BP; 255 A; 353 C; 331 G; 252 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 7; Length 1191;  
 Best Local Similarity 90.9%; Pred. No. 2.2e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 CGGCGGCGAGCCAAACCCAGGAT 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 328 CGGCGGCGGAGCAAAACCCAGGAT 307

RESULT 12  
 ABV78157/c  
 ID ABV78157 standard; DNA; 486 BP.  
 XX  
 AC ABV78157;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human ID4 DNA SEQ ID NO 41.  
 XX  
 KW RNA inhibition; dsRNA; gene expression inhibitor; oncogene; cytostatic;  
 KW virucide; protozoacide; gene; ds.  
 XX  
 OS *Homo sapiens*.  
 XX  
 PN W0200255693-A2.  
 XX  
 PD 18-JUL-2002.  
 XX  
 XX 09-JAN-2002; 2002WO-EP000152.  
 PF  
 XX 09-JAN-2001; 2001DE-01000586.  
 PR  
 PR 26-OCT-2001; 2001DE-01055280.  
 PR  
 PR 29-NOV-2001; 2001DE-01058411.  
 PR  
 PR 07-DEC-2001; 2001DE-01060151.  
 XX  
 XX (RIBO-) RIBOPHARMA AG.  
 PA  
 XX Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 XX  
 XX WPI; 2002-590671/63.

XX Inhibiting expression of target gene, useful e.g. for inhibiting  
PT oncogenes, by administering double-stranded RNA complementary to the  
PT target and having an overhang.  
XX  
XX  
PS Claim 10; Page 136; 203pp; German.  
XX  
CC The invention relates to inhibiting expression of a target gene (I) in a  
CC cell by introducing an inhibitory RNA (dsRNA) having a double-stranded  
CC structure of at most 49 consecutive bases. At least part of one strand  
CC (as1) of dsRNA is complementary to (I) and at least one end of dsRNA  
CC has an overhang of 1-4 nucleotides. The method is used to inhibit the  
CC expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.  
CC in humans, also genes in plasmidium or in viruses or viroids that are  
CC pathogenic for humans, animals or plants. Introducing an overhang into  
CC dsRNA greatly increases effectiveness for inhibiting gene expression,  
CC both in vivo and in vitro and also increases stability and thus the  
CC effective concentration inside the cell. The present sequence is that of  
CC a gene related to the invention  
XX  
SQ Sequence 486 BP; 77 A; 181 C; 174 G; 54 T; 0 U; 0 Other;  
  
Query Match 74.4%; Score 18.6; DB 6; Length 486;  
Best Local Similarity 84.0%; Pred. No. 2.5e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CCGCGGGCGGAGAGCCCAACCCAGGAT 25  
DB 124 CCGCGGCTGCGGAGCCCAACCCAGGCT 100  
  
RESULT 13  
ABZ35733/C  
ID ABZ35733 standard; DNA; 486 BP.  
AC ABZ35733;  
XX  
DT 07-FEB-2003 (first entry)  
DE Human ID4 polynucleotide SEQ ID NO 41.  
XX  
XX Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;  
KW protozoicide; gene expression; antisense; tumour; infection; Plasmidium;  
KW virus; viroid; anti-GFP; human; HIV; human immunodeficiency virus;  
KW Hepatitis C virus; human papilloma virus; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN DE10100588-A1.  
XX  
PD 18-JUL-2002.  
XX  
PF 09-JAN-2001; 2001DE-01000588.  
XX  
PR 09-JAN-2001; 2001DE-01000588.  
XX  
PA (RIBO-) RIBOPHARMA AG.  
XX  
PI Kreutzer R, Limmer S, Rost S, Hadwiger P;  
XX  
DR WPI; 2002-683450/74.  
XX  
XX Inhibiting expression of target genes, useful e.g. for treating tumors,  
PT by introducing into cells two double-stranded RNAs that are complementary  
PT to the target.  
XX  
XX Claim 13; Page 32; 100pp; German.  
XX  
CC The invention relates to inhibiting expression of a target gene in a cell  
CC by introducing at least two oligonucleotides (dsRNA and II), both  
CC with a double-stranded (ds) structure of at most 49 sequential nucleotide  
CC pairs. At least part of one strand (S1, S2) of the ds structures in each  
CC of dsRNAI and II are complementary to regions in the target gene. The

CC method uses antisense inhibition of gene expression using double stranded  
CC RNA inhibition (RNAI). The method is particularly used to treat tumours  
CC or infections, especially by plasmidium or viruses/viroids (pathogenic on  
CC humans, animals or plants). The method provides more effective inhibition  
CC of expression than known methods using a single dsRNA, even at very low  
CC concentrations. When dsRNA has at least one unpaired nucleotide at the  
CC end, stability (and thus effective concentration in the cell) is improved  
CC and efficiency can be increased further by pretreating the cells with  
CC interferon. The present sequence is that of a target DNA of the invention  
XX  
SQ Sequence 486 BP; 77 A; 181 C; 174 G; 54 T; 0 U; 0 Other;  
  
Query Match 74.4%; Score 18.6; DB 6; Length 486;  
Best Local Similarity 84.0%; Pred. No. 2.5e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CCGCGGGCGGAGAGCCCAACCCAGGAT 25  
DB 124 CCGCGGCTGCGGAGCCCAACCCAGGCT 100  
  
RESULT 14  
ABX09976/C  
ID ABX09976 standard; DNA; 486 BP.  
XX  
AC ABX09976;  
XX  
DT 23-JAN-2003 (first entry)  
XX  
DE Human ID4 DNA fragment SEQ ID 41.  
XX  
XX Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;  
KW prion; inhibition; human; ds.  
XX  
OS Homo sapiens.  
XX  
PN DE10100587-C1.  
XX  
PD 21-NOV-2002.  
XX  
PF 09-JAN-2001; 2001DE-01000587.  
XX  
PR 09-JAN-2001; 2001DE-01000587.  
XX  
PA (RIBO-) RIBOPHARMA AG.  
XX  
PI Kreutzer R, Limmer S, Rost S, Hadwiger P;  
XX  
DR WPI; 2002-742209/81.  
XX  
XX Inhibiting expression of target genes, e.g. oncogenes, in cells, by  
PT introduction of complementary double-stranded oligoribonucleotide, after  
PT treating the cell with interferon.  
XX  
PS Disclosure; Page 37; 98pp; German.  
XX  
XX This invention describes a novel method for inhibiting expression of a  
CC target gene by introducing into the cell that contains the target gene at  
CC least one oligoribonucleotide (dsRNAI) that has a double-stranded (ds)  
CC structure of not more than 49 consecutive nucleotides (nt), where at  
CC least a segment of one strand of the ds structure is complementary with  
CC the target gene and the cells are treated with interferon before  
CC introduction of dsRNAI. The method is used to inhibit expression of  
CC target genes, particularly oncogenes, cytokine genes, Id (not defined)  
CC protein genes; developmental or prion genes, or genes expressed in  
CC pathogenic organisms (particularly plasmidia) or in viruses or viroids  
CC (pathogenic in humans, animals or plants). Treating the cells with  
CC interferon greatly increases the extent to which dsRNA can inhibit  
CC expression of the target genes, and the effect is even greater when dsRNA  
CC are modified to increase their stability. ABX09976-ABX10075 represent  
CC gene fragments used to illustrate the method of the invention  
XX  
XX Hence 486 BP; 77 A; 181 C; 174 G; 54 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 486;  
 Best Local Similarity 84.0%; Pred. No. 2.5e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGCGCGCGGAGAGCCACCCAGGAT 25  
 |||||  
 Db 124 CCGCGGCTGCGGAGCCACCCAGGCT 100

## RESULT 15

ABL91698/c  
 ID ABL91698 standard; DNA; 486 BP.

XX AC ABL91698;

XX DT 28-MAY-2002 (first entry)

XX DE Human polynucleotide SEQ ID NO 41.

XX KW Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;  
 KW Plasmodium; virus; viroid; cytokine; prion; antisense oligonucleotide;  
 KW cytosstatic; virucide; protozoacide; antibacterial; ds.

XX OS Homo sapiens.

XX PN DE10100586-Cl.

XX PD 11-APR-2002.

XX PF 09-JAN-2001; 2001DE-01000586.

XX PR 09-JAN-2001; 2001DE-01000586.

XX PA (RIBO-) RIBOPHARMA AG.

XX PI Kreutzer R, Limmer S, Rost S, Hadwiger P;

XX DR WPI; 2002-270454/32.

XX PT Inhibiting gene expression in cells, useful for e.g. treating tumors, by  
 PT introducing double-stranded complementary oligorNA having unpaired  
 PT terminal bases.

XX PS Claim 13; Page 34; 104pp; German.

XX CC The invention relates to a method for inhibiting expression of a target  
 CC gene (ABL91658-ABL91797) in a cell by introducing at least one  
 CC oligoribonucleotide that has a double-stranded structure consisting of at  
 CC most 49 sequential nucleotide pairs, with at least part of one strand  
 CC complementary with the target gene and has at least one end a single-  
 CC stranded segment of 1-4 nt. The method provides oligoribonucleotides for  
 CC antisense inhibition of gene expression useful e.g. for treating tumors  
 CC but the oligoribonucleotides may also be directed against genes present  
 CC in pathogens (e.g. Plasmodium or viruses/viroids, pathogenic on humans,  
 CC animals or plants) or against cytokine, ID, developmental or prion genes.  
 CC The method provides more effective inhibition of gene expression than use  
 CC of known oligonucleotides, probably because the unpaired overhang  
 CC increases stability and thus intracellular concentration.

XX SQ Sequence 486 BP; 77 A; 181 C; 174 G; 54 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 486;  
 Best Local Similarity 84.0%; Pred. No. 2.5e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGCGCGCGGAGAGCCACCCAGGAT 25  
 |||||  
 Db 124 CCGCGGCTGCGGAGCCACCCAGGCT 100

Search completed: April 29, 2004, 04:57:15  
 Job time : 209.473 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 206.373 Seconds  
(without alignments)  
514.627 Million cell updates/sec

Title: US-10-624-714-10

Perfect score: 25

Sequence: 1 ccgcggcgccagagccaccaggat 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	110000	4	AAI99682_11
C 2	25	100.0	110000	4	AAI99683_11
C 3	19.8	79.2	2000	7	ADA73398
C 4	19.2	76.8	1110	3	AAZ53286
C 5	19.2	76.8	1110	3	AAZ53285
C 6	19.2	76.8	102634	3	AAA81464
C 7	19.2	76.8	110000	3	AAA81490_10
C 8	19.2	76.8	349980	3	AAZ1609
C 9	18.8	75.2	889	4	AAH02144
C 10	18.8	75.2	918	7	ACA43747
C 11	18.8	75.2	1191	7	ACA43762
C 12	18.6	74.4	486	6	ABV78157
C 13	18.6	74.4	486	6	ABZ35733
C 14	18.6	74.4	486	6	ABX09976
C 15	18.6	74.4	486	6	ABL91698
C 16	18.6	74.4	1322	4	ABD11133
C 17	18.6	74.4	3907	6	ABQ54325
C 18	18.6	74.4	4813	7	ABT16935
C 19	18.6	74.4	4813	7	ABZ67594
C 20	18.6	74.4	4813	9	ADC20738
C 21	18.6	74.4	185371	6	ABT10718
C 22	18.2	72.8	731	5	AAS66782
C 23	18.2	72.8	1026	4	AAS4064

24	18.2	72.8	1026	7	ACA42146	ACA42146	Prokaryot
25	18.2	72.8	5236	4	AAK51656	AAK51656	Human pol
26	18.2	72.8	5530	4	AAK52640	AAK52640	Human pol
27	18.2	72.8	7769	6	ABN96867	ABN96867	Gene #336
C 28	18.2	72.8	53178	4	AAAS5343	AAAS5343	Propionib
C 29	18.2	72.8	53178	7	ACF64472	ACF64472	Propionib
C 30	17.8	71.2	793	5	AAAS71344	AAAS71344	DNA encod
C 31	17.8	71.2	906	5	AAAS6586	AAAS6586	DNA encod
C 32	17.8	71.2	4647	6	ABA90426	ABA90426	Drosophil
C 33	17.8	71.2	5316	4	ABL11843	ABL11843	Drosophil
C 34	17.8	71.2	7499	4	ABL11842	ABL11842	Drosophil
C 35	17.8	71.2	135638	7	ABX34289	ABX34289	S. atrop
C 36	17.6	70.4	162	3	AAAS2038	AAAS2038	N. mening
C 37	17.6	70.4	711	4	AAH07034	AAH07034	Human cDN
C 38	17.6	70.4	733	4	AAH03810	AAH03810	Human cDN
C 39	17.6	70.4	756	4	AAH03708	AAH03708	Human cDN
C 40	17.6	70.4	873	4	AAH03186	AAH03186	Human cDN
C 41	17.6	70.4	885	7	ACA26953	ACA26953	Prokaryot
C 42	17.6	70.4	1006	8	ACF08242	ACF08242	Human NOV
C 43	17.6	70.4	1216	7	ACD19359	ACD19359	CDNA enco
C 44	17.6	70.4	1216	7	ACD19360	ACD19360	CDNA enco
C 45	17.6	70.4	1216	7	ACD19361	ACD19361	CDNA enco

## ALIGNMENTS

## RESULT 1

AAI99682\_11/c  
Continuation (12 of 45) of AAI99682 from base 1100001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000

WP AAI99682\_41 410001 4210000  
WP AAI99682\_42 420001 4310000  
WP AAI99682\_43 430001 4410000  
WP AAI99682\_44 440001 4411529

Query Match 100.0%; Score 25; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.69; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGGGGCGAGCCCAACCGGAT 25  
|||  
Db 13135 CCGCGGGGCGAGCCCAACCGGAT 13111

RESULT 2  
AAI99683\_11/c  
Continuation (12 of 44) of AAI99683 from base 1100001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683

WP Fragment Name Begin End  
WP AAI99683\_00 1 110000  
WP AAI99683\_01 10001 210000  
WP AAI99683\_02 20001 310000  
WP AAI99683\_03 30001 410000  
WP AAI99683\_04 40001 510000  
WP AAI99683\_05 50001 610000  
WP AAI99683\_06 60001 710000  
WP AAI99683\_07 70001 810000  
WP AAI99683\_08 80001 910000  
WP AAI99683\_09 90001 1010000  
WP AAI99683\_10 100001 110000  
WP AAI99683\_11 110001 1210000  
WP AAI99683\_12 120001 1310000  
WP AAI99683\_13 130001 1410000  
WP AAI99683\_14 140001 1510000  
WP AAI99683\_15 150001 1610000  
WP AAI99683\_16 160001 1710000  
WP AAI99683\_17 170001 1810000  
WP AAI99683\_18 180001 1910000  
WP AAI99683\_19 190001 2010000  
WP AAI99683\_20 200001 2110000  
WP AAI99683\_21 210001 2210000  
WP AAI99683\_22 220001 2310000  
WP AAI99683\_23 230001 2410000  
WP AAI99683\_24 240001 2510000  
WP AAI99683\_25 250001 2610000  
WP AAI99683\_26 260001 2710000  
WP AAI99683\_27 270001 2810000  
WP AAI99683\_28 280001 2910000  
WP AAI99683\_29 290001 3010000  
WP AAI99683\_30 300001 3110000  
WP AAI99683\_31 310001 3210000  
WP AAI99683\_32 320001 3310000  
WP AAI99683\_33 330001 3410000  
WP AAI99683\_34 340001 3510000  
WP AAI99683\_35 350001 3610000  
WP AAI99683\_36 360001 3710000  
WP AAI99683\_37 370001 3810000  
WP AAI99683\_38 380001 3910000  
WP AAI99683\_39 390001 4010000  
WP AAI99683\_40 400001 4110000  
WP AAI99683\_41 410001 4210000  
WP AAI99683\_42 420001 4310000  
WP AAI99683\_43 430001 4403765

Query Match 100.0%; Score 25; DB 4; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.69; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGGGGCGAGCCCAACCGGAT 25  
|||  
Db 13158 CCGCGGGGCGAGCCCAACCGGAT 13134

RESULT 3  
ADA73398  
ID ADA73398 standard; DNA; 2000 BP.

XX AC ADA73398;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 6724.

XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
gene expression.

PS Claim 27; SEQ ID NO 6724; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
involved in plant resistance or response to pathogenic infection. M1  
comprises identifying a gene whose expression is significantly altered in  
the incompatible interaction of plant gene expression relative to  
expression of the gene in an uninfected plant, in a mutant plant that  
does not express a gene associated with response to pathogenic infection,  
or in a corresponding incompatible or compatible interaction. (M1) is  
useful for conferring resistance to resistance or tolerance to a plant to  
bacterial, fungal or viral infection. The present sequence was used to  
illustrate the invention.

SQ Sequence 2000 BP; 502 A; 434 C; 481 G; 582 T; 0 U; 1 Other;

Query Match 79.2%; Score 19.9; DB 7; Length 2000;

Best Local Similarity 91.3%; Pred. No. 85;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGGGGCGAGCCCAACCGGAT 23  
|||

Db 1862 CCGCGGGGCGAGCCCAACCGGAT 1884

RESULT 4

AAZ53286/c

ID AAZ53286 standard; DNA; 1110 BP.

XX AC AAZ53286;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria meningitidis ORF 135 partial DNA sequence SEQ ID NO:521.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy; ds.

XX

OS Neisseria meningitidis.  
 XX WO9957280-A2.  
 PN PD 11-NOV-1999.  
 XX PF 30-APR-1999; 99WO-US009346.  
 XX PR 01-MAY-1998; 98US-0083758P.  
 XX PR 31-JUL-1998; 98US-0094869P.  
 XX PR 02-SEP-1998; 98US-0098994P.  
 XX PR 02-SEP-1998; 98US-0099062P.  
 XX PR 09-OCT-1998; 98US-0103794P.  
 XX PR 09-OCT-1998; 98US-0103794P.  
 XX PR 09-OCT-1998; 98US-0103796P.  
 XX PR 25-FEB-1999; 99US-0121528P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX DR WPI: 2000-062150/05.  
 XX DR P-PSDB; AAY74524.  
 XX PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics.  
 XX PS Claim 7; Page 384; 1453pp; English.  
 XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitis and N. gonorrhea polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of the  
 CC invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the manufacture  
 CC of medicaments for treating or preventing infection due to Neisserial  
 CC bacteria (e.g. meningitis and septicemia), to detect the presence of  
 CC Neisseria bacteria, or to raise antibodies. They may also be used to  
 CC screen for agonists or antagonists, which may themselves have use as  
 CC antibacterial agents. The polynucleotides of the invention may also be  
 CC used in gene therapy protocols  
 XX SQ Sequence 1110 BP; 268 A; 314 C; 309 G; 219 T; 0 U; 0 Other;  
 Query Match 76.8%; Score 19.2; DB 3; Length 1110;  
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 CCGCGCGCGCAGCCGACACGGA 24  
 DB 1009 CGCGCGCGCAGCGCCGACGGA 986  
 RESULT 5  
 AAZ53285/C  
 ID AAZ53285 standard; DNA; 1110 BP.  
 XX AC AAZ53285;  
 XX DT 21-MAR-2000 (first entry)  
 XX DE Neisseria meningitidis ORF 135 partial DNA sequence SEQ ID NO:519.  
 XX KW Neisseria meningitidis; Neisseria gonorrhea; antigen; vaccine;  
 XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 XX KW antibacterial; gene therapy; ds.  
 XX OS Neisseria meningitidis.  
 XX PN WO200022430-A2.

PN WO9957280-A2.  
 XX PD 11-NOV-1999.  
 XX PF 30-APR-1999; 99WO-US009346.  
 XX PR 01-MAY-1998; 98US-0083758P.  
 XX PR 31-JUL-1998; 98US-0094869P.  
 XX PR 02-SEP-1998; 98US-0098994P.  
 XX PR 02-SEP-1998; 98US-0099062P.  
 XX PR 09-OCT-1998; 98US-0103794P.  
 XX PR 09-OCT-1998; 98US-0103794P.  
 XX PR 09-OCT-1998; 98US-0103796P.  
 XX PR 25-FEB-1999; 99US-0121528P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX DR WPI: 2000-062150/05.  
 XX DR P-PSDB; AAY74523.  
 XX PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics.  
 XX PS Claim 7; Page 382-383; 1453pp; English.  
 XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitis and N. gonorrhea polynucleotides  
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of the  
 CC invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the manufacture  
 CC of medicaments for treating or preventing infection due to Neisserial  
 CC bacteria (e.g. meningitis and septicemia), to detect the presence of  
 CC Neisseria bacteria, or to raise antibodies. They may also be used to  
 CC screen for agonists or antagonists, which may themselves have use as  
 CC antibacterial agents. The polynucleotides of the invention may also be  
 CC used in gene therapy protocols  
 XX SQ Sequence 1110 BP; 264 A; 322 C; 309 G; 215 T; 0 U; 0 Other;  
 Query Match 76.8%; Score 19.2; DB 3; Length 1110;  
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 CCGCGCGCGCAGCCGACACGGA 24  
 DB 1009 CGCGCGCGCAGCGCCGACGGA 986  
 RESULT 6  
 AA81464  
 ID AA81464 standard; DNA; 102634 BP.  
 XX AC AA81464;  
 XX DT 04-DEC-2000 (first entry)  
 XX DE N. meningitidis partial DNA sequence gnm\_12 SEQ ID NO:12.  
 XX KW Neisseria meningitidis; Neisseria gonorrhea; genome; immunogenic;  
 XX KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 XX KW Meningococcus B; MenB; ds.  
 XX OS Neisseria meningitidis.  
 XX PN WO200022430-A2.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 466.667 Seconds  
(without alignments)  
242.328 Million cell updates/sec

Title: US-10-624-714-10

Perfect score: 25

Sequence: 1 ccgcggcgagagcaaccaggat 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	75.2	918	13	US-10-282-122A-31617
C 2	18.8	75.2	1191	13	US-10-282-122A-31632
C 3	18.6	74.4	816	15	US-10-023-386-22934
C 4	18.6	74.4	3907	16	US-10-264-049-205
C 5	18.4	73.6	488	16	US-10-260-238-394
C 6	18.4	73.6	145025	13	US-10-087-192-1051
C 7	18.2	72.8	1026	9	US-09-815-243-7701
C 8	18.2	72.8	1026	13	US-10-282-122A-30016
C 9	17.8	71.2	108	13	US-10-085-783A-6242
C 10	17.8	71.2	108	16	US-10-242-355A-6242
C 11	17.8	71.2	4847	15	US-10-161-051-75
C 12	17.8	71.2	135638	15	US-10-314-657-1
C 13	17.6	70.4	613	13	US-10-425-114-22102
C 14	17.6	70.4	885	13	US-10-282-122A-14823

15 17.6 70.4 912 13 US-10-425-114-21673 Sequence 21673, A

16 17.6 70.4 942 15 US-10-156-761-2483 Sequence 2483, Ap

17 17.6 70.4 979 16 US-10-369-493-27740 Sequence 27740, A

18 17.6 70.4 1006 16 US-10-307-928A-19 Sequence 19, Appl

19 17.6 70.4 1062 13 US-10-425-114-17508 Sequence 17508, A

20 17.6 70.4 1260 15 US-10-156-761-7420 Sequence 7420, Ap

21 17.6 70.4 1363 9 US-09-880-192-27 Sequence 27, Appl

22 17.6 70.4 1363 15 US-10-427-348-27 Sequence 27, Appl

23 17.6 70.4 1494 9 US-09-728-952-80 Sequence 80, Appl

24 17.6 70.4 1628 16 US-10-398-221-1718 Sequence 1718, Ap

25 17.6 70.4 1732 16 US-10-369-493-27128 Sequence 27128, A

26 17.6 70.4 2074 16 US-10-108-260A-980 Sequence 980, App

27 17.6 70.4 2115 15 US-10-154-386-3 Sequence 3, Appl

28 17.6 70.4 2223 15 US-10-154-386-1 Sequence 1, Appl

29 17.6 70.4 2596 16 US-10-108-260A-1271 Sequence 1271, Ap

30 17.6 70.4 10851 10 US-09-888-233A-2 Sequence 2, Appl

31 17.6 70.4 9025608 15 US-10-156-761-1 Sequence 1, Appl

32 17.4 946 13 US-10-282-122A-23850 Sequence 23850, A

33 17.2 68.8 256 13 US-10-424-599-47813 Sequence 47813, A

34 17.2 68.8 450 13 US-10-027-632-274644 Sequence 274644, A

35 17.2 68.8 450 13 US-10-027-632-274645 Sequence 274645, A

36 17.2 68.8 450 13 US-10-027-632-274647 Sequence 274647, A

37 17.2 68.8 450 16 US-10-027-632-274644 Sequence 274644, A

38 17.2 68.8 450 16 US-10-027-632-274645 Sequence 274645, A

39 17.2 68.8 450 16 US-10-027-632-274647 Sequence 274647, A

40 17.2 68.8 467 10 US-09-918-995-2694 Sequence 2694, Ap

41 17.2 68.8 723 13 US-10-425-114-6582 Sequence 5882, Ap

42 17.2 68.8 748 13 US-10-027-632-25652 Sequence 25652, A

43 17.2 68.8 748 16 US-10-027-632-25652 Sequence 147, App

44 17.2 68.8 897 10 US-09-989-643-147 Sequence 23137, A

45 17.2 68.8 975 13 US-10-282-122A-23137 Sequence 23137, A

#### ALIGNMENTS

#### RESULT 1

US-10-282-122A-31617/c  
; Sequence 31617, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282.122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22



Db 491 CCGCGGCTGGGAGGCCACCCAGGCT 467  
|||||

RESULT 5  
US-10-260-238-394  
; Sequence 394, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Rumiya  
; APPLICANT: Krepis, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 394  
; LENGTH: 488  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
US-10-260-238-394

Query Match 73.8%; Score 18.4; DB 16; Length 488;  
Best Local Similarity 95.0%; Pred. No. 93;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGCGGCGGAGGCCAACC 20  
|||||

Db 317 CCGCGGCGGAGGCCATCC 336  
|||||

RESULT 6  
US-10-087-192-1051  
; Sequence 1051, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: CANCER  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1051  
; LENGTH: 145025  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...((145025))  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-1051

Query Match 73.8%; Score 18.4; DB 13; Length 145025;  
Best Local Similarity 95.0%; Pred. No. 51;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 5 GCGCGGAGGCCAACCAGGA 24  
|||||

Db 14160 GCGCGGAGGCCATCCAGGA 14179  
|||||

RESULT 7  
US-09-815-242-7701  
; Sequence 7701, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7701  
; LENGTH: 1026  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1026)  
US-09-815-242-7701

Query Match 72.8%; Score 18.2; DB 9; Length 1026;  
Best Local Similarity 87.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGCGGCGGAGGCCAACCAGG 23  
|||||

Db 810 CCGCGGCGGTCAGCGCCAACCAGG 832  
|||||

RESULT 8  
US-10-282-122A-30016  
; Sequence 30016, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
US-10-282-122A-30016





```
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (77)..(77)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87)..(87)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-6242

Query Match      71.2%; Score 17.8; DB 16; Length 108;
Best Local Similarity 82.6%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCGGCGGCGAGCCACACAGGAT 25
Db 55 GCGGCGGCGAGGAGANACACAGGNT 33

RESULT 11
US-10-161-051-75
; Sequence 75, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Midgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 4647
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-161-051-75

Query Match      71.2%; Score 17.8; DB 15; Length 4647;
Best Local Similarity 90.5%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGGCGGCGAGCCACACAGG 23
Db 3671 GCGGCGGCGAGTGCACACCAAG 3691

RESULT 12
US-10-314-657-1/c
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA

; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (77)..(77)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87)..(87)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-6242

Query Match      71.2%; Score 17.8; DB 15; Length 135638;
Best Local Similarity 90.5%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGGCGGCGAGCCACACAGG 23
Db 79524 GCGGCGGCGAGCCCGCCAGG 79504

RESULT 13
US-10-425-114-22102/c
; Sequence 22102, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22102
; LENGTH: 613
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3356-051-D9_FLI
US-10-425-114-22102

Query Match      70.4%; Score 17.6; DB 13; Length 613;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGCGGCGAGCCACACAGGA 24
Db 240 CCGGCGGCGAGCCACACAGGA 217

RESULT 14
US-10-282-122A-14823
; Sequence 14823, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```

, PRIOR APPLICATION NUMBER: 60/230,333
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/230,347
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/242,578
, PRIOR FILING DATE: 2000-10-23
, PRIOR APPLICATION NUMBER: 60/253,625
, PRIOR FILING DATE: 2000-11-27
, PRIOR APPLICATION NUMBER: 60/257,931
, PRIOR FILING DATE: 2000-12-22
, PRIOR APPLICATION NUMBER: 60/267,636
, PRIOR FILING DATE: 2001-02-09
, PRIOR APPLICATION NUMBER: 60/269,308
, PRIOR FILING DATE: 2001-02-16
, Remaining Prior Application data removed
, NUMBER OF SEQ ID NOS: 78614
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 14823
, LENGTH: 885
, TYPE: DNA
, ORGANISM: Bordetella pertussis
US-10-282-132A-14823

```

	Query Match	70.4%	Score 17.6;	DB 13;	Length 885;
	Best local Similarity	83.3%;	Prod. NO. 2e+02;		
	Matches 20;	Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;
Oy	1	CCGCGCGCGGAGACCAACGGA	24		
Db	762	CCGCGCGGAGAGCGCCACACGA	785		

```

RESULT 15
US-10-425-114-21673
; Sequence 21673, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kowalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF-SEQ ID NOS: 73128
; SEQ ID NO 21673
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-070-A3_FLI
US-10-425-114-21673

```

```

Query Match          70.4%; Score 17.6; DB 13; Length 912;
Best Local Similarity 83.3%; Prid. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  CCGCGCGCGCAGAGCCAAACGAGCA 24
          |||||
DB       200 CCGGTGCAGCAGAGCCAAACCAAGA 223

```

Search completed: April 29, 2004, 20:45:07  
Job time : 471.667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 377.255 Seconds  
(without alignments)  
2872.264 Million cell updates/sec

Title: US-10-624-714-10

Perfect score: 25  
Sequence: 1 ccgcgcgcgcagccaccaggat 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.ste.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.nam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	25	100.0	20916	1	AE006986	AE006986 Mycobacte
C 2	25	100.0	327650	1	BX248337	BX248337 Mycobacte
C 3	25	100.0	349306	15	BX842575	BX842575 Mycobacte
C 4	21.8	87.2	37304	1	MLCL373	AL035500 Mycobacte
C 5	21.8	87.2	344050	1	MLEPRTN1	AL035500 Mycobacte
C 6	20.4	81.6	1309	8	AF063800	AF063800 Odontella
C 7	20.2	80.8	732	9	HA333853	AJ333853 Homo sapi
C 8	20.2	80.8	346274	1	BX640430	BX640430 Bordetell
C 9	20.2	80.8	348014	1	BX640430	BX640430 Bordetell
C 10	20.2	80.8	349672	1	BX640419	BX640419 Bordetell
C 11	19.8	79.2	2000	6	AX658854	AX658854 Sequence
C 12	19.8	79.2	162425	8	AC107226	AC107226 Oryza sat
C 13	19.8	79.2	218467	2	AC135355	AC135355 Mus muscu
C 14	19.8	79.2	226332	2	AC145568	AC145568 Mus muscu
C 15	19.8	79.2	231638	2	AC123190	AC123190 Rattus no
C 16	19.2	76.8	1981	3	AF002806	AF002806 Leistuss f
C 17	19.2	76.8	10040	1	AE002457	AE002457 Neisseria
C 18	19.2	76.8	220663	2	AC094574	AC094574 Rattus no
C 19	19.2	76.8	237893	2	AC094418	AC094418 Rattus no
C 20	19.2	76.8	301200	1	AF005374	AF005374 Thermosyn
C 21	19.2	76.8	331801	1	NMA422491	AL162755 Neisseria
C 22	19.2	76.8	349980	6	AX044031	AX044031 Sequence
C 23	18.8	75.2	889	6	AX111404	AX111404 Sequence
C 24	18.8	75.2	168837	2	AC145464	AC145464 Oryctolag
C 25	18.8	75.2	300511	1	AE016775	AE016775 Pseudomon
C 26	18.6	74.4	486	6	AX481427	AX481427 Sequence
C 27	18.6	74.4	1147	9	HSID4	Y07958 H.sapiens m
C 28	18.6	74.4	1322	6	AX201585	AX201585 Sequence
C 29	18.6	74.4	1322	9	HSU28368	U98368 Human Id-re
C 30	18.6	74.4	2389	9	BC014941	BC014941 Homo sapi
C 31	18.6	74.4	128379	9	HSJ773A18	AL049557 Human DNA
C 32	18.6	74.4	139490	8	AF003225	AF003225 Oryza sat
C 33	18.6	74.4	149995	9	AC016722	AC016722 Homo sapi
C 34	18.6	74.4	160114	2	AC022181	AC022181 Homo sapi
C 35	18.6	74.4	185371	9	HS625H18	AL022726 Human DNA
C 36	18.4	73.6	49730	2	AC101650	AC101650 Mus muscu
C 37	18.4	73.6	114997	10	AL954374	AL954374 Mouse DNA
C 38	18.4	73.6	131387	10	AC079440	AC079440 Mus muscu
C 39	18.4	73.6	208065	2	AC107757	AC107757 Mus muscu
C 40	18.4	73.6	244594	2	AC112470	AC112470 Rattus no
C 41	18.4	73.6	245560	1	AF005330	AF005330 Vibrio vu
C 42	18.2	72.8	5241	9	AK125717	AK125717 Homo sapi
C 43	18.2	72.8	6499	9	HSMIGST01	U71210 Homo sapien
C 44	18.2	72.8	7769	6	AX410718	AX410718 Sequence
C 45	18.2	72.8	7769	9	AH008642S1	AF092926 Homo sapi

# ALIGNMENTS

RESULT 1  
AE006986/c  
LOCUS AE006986 20916 bp DNA linear BCT 27-APR-2001  
DEFINITION Mycobacterium tuberculosis CDC1551, section 72 of 280 of the complete genome.  
ACCESSION AE006986 AE000516  
VERSION AE006986.1 GI:13880583  
KEYWORDS Mycobacterium tuberculosis CDC1551  
SOURCE Mycobacterium tuberculosis CDC1551  
ORGANISM Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.  
1 (bases 1 to 20916)  
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 20916)  
Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. .20916  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
complement (95. .1096)  
/gene="MT1018"  
complement (95. .1096)  
/gene="MT1018"  
/note="similar to SP:P55785 PID:1731683; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="poly(phenyl synthase"  
/protein\_id="AAK45265.1"  
/db\_xref="GI:13880584"  
/translation="MRMTATSMIPAVSLGSDPQFTANVHDGIARTILNLSLQADE VMDTVHLVDAGCTPFLFTVLAALQSLDPDGEVTVAGAAIELMHGTHCHDRVV DESDMRGKTPSDNTNNPFAILAGDYRFATASQLASRLDPFAFAVFAEFLITGQ MRATRGPAISHDITIEHLRVVHEKTSGLIAASQALGAASGAABEIRRVARLGRMIG APEISRDIIAISGDSATLSGADLGOAVHTPLMYALREQTPTSLRELLAGPTHDD HVABATLLRCSGIGKAKNVVAAQAAREELPYLPDRQPRALATLIDHVASACD" complement (1133. .1789)  
/gene="MT1019"  
complement (1133. .1789)  
/gene="MT1019"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45266.1"  
/db\_xref="GI:13880585"  
/translation="MAESSLNPSLSVRISAFILRPDWTWTRVRAARRFAAAGLVLAGVAA LRNPEDREVVAAHDLRPGTALTPGVDRLEKRSATLPPGSAADLDAVVGSTIAS PTRRGEVLTVRLGSLRAESTAGPDARTVPLHLADSLVLRVGVGVDDVLAAPVTD SPAALRLATDAIVLVSQAQQAADSRLVVALPARLANTVAGAAALGQTVTLTH" complement (1862. .2215)  
/gene="MT1020"  
complement (1862. .2215)  
/gene="MT1020"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45267.1"  
/db\_xref="GI:13880586"  
/translation="MTFPEVLVPTYSVECTQCANREDVVQAFTDDALTTCCSGRLR KLFNAVGVFKGTFYRTDRSGKSKSQTSQTSSTSESTKSGSGSGSGSGSKASG STEKSSSTTAAAV" complement (2268. .2855)  
/gene="MT1021"  
complement (2268. .2855)  
/gene="MT1021"  
/note="similar to GP:4455709; identified by sequence

similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="5-formyltetrahydrofolate cyclo-ligase-related protein"  
/protein\_id="AAK45268.1"  
/db\_xref="GI:13880587"  
/translation="MASKSLRDQLLAARBRVADDVRAAEARMLRGHLERMTVSDSTV CAIVPGEGFGESIEMLDVLRRAGRVLLPVARTAGDGLPLPLPWGVRAGGLARAWG LLEFPBPWLEAALAAQASLVLPALAVDROGVSLGRGFGYDRSLKCRDHPARLVAVV RTVELVDVLPSEPHDVEPMTHALTFFERGLIALPCGE" 2962. .3882  
/gene="MT1022"  
2962. .3882  
/gene="MT1022"  
/note="similar to GB:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UTP--glucose-1-phosphate uridylyltransferase"  
/protein\_id="AAK45269.1"  
/db\_xref="GI:13880588"  
/translation="MSRPEVLTFTAIIVPAAGLGRPLPATKVPKELLPVVTPGIE LVAAEAAAGAEELVIVTSGKGVVAHFVEDLVLEGTLEARGKIAMLAKVRAPALI KVESVQAEPFLGLGHAIGCVPEPLSPDGDVAVLLPDLVPLTGVLETSKVRASGG TVLCAIEVAREEISAYGVFDVPEVPDGDYDDPNVLKVRGMVEKPKAETAPSYAAAG RYVLDRAIFDRLRRI DRGAGEVQLTDAIALIAEGHPVHVHVQSGSRHDLGNPGGYL KAAVDFAIDRDDYGPDLRLRLVLAELGITEQ" 3959. .5239  
/gene="MT1023"  
3959. .5239  
/gene="MT1023"  
/note="similar to GB:AE000666; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="molybdopterin biosynthesis protein MoaA"  
/protein\_id="AAK45270.1"  
/db\_xref="GI:13880589"  
/translation="MRSVEEQQAISAAAVAPRIRVAIAEQALMCAEEVWTPMP GFDCAAIDGVAVRSVDVAGVDTGVGFADHGDLDGRDVLTLPMVTIEAGARTLSR LQPRQAVRVTGAPLPTLADAVLPLWTDGMSRVRVLRGAPSGAVYVRAGDDVQPD VAPRAGTIIIGAAQVGLLAAGVRERVLVHPRRLSVAVAGSELVISTPCNGQVYDVN SYALAAAGRDCAEVRNIVSNDPTSELGEIVFPGQLNRAEVVVIAGVGGAAAEVRS VLSLGEVVRVAVHFGVQVQGLGRDGVPTLLPANPVSAVLFVFEVVRPLRLS LGKHPMRIRVARTLSPITSVAGKGYLRGQMLRQDQSGEYLVAQLGGAGASSHLL ATLAENCLVVPVPTGAEQIRTGTEIVDVAFLAQHG" 5302. .5913  
/gene="MT1024"  
5302. .5913  
/gene="MT1024"  
/note="similar to GB:M99278 SP:P09454 GB:X06118 PID:147655 PID:142744; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal-protein-alanine acetyltransferase"  
/protein\_id="AAK45271.1"  
/db\_xref="GI:13880590"  
/translation="MAVGPLRVSAQVIRLRPVRMDGVHWSRIRLADRAHLEPWPESA DGEWTVRHTVAAMPVAVCSGLRSEARNGMLFYIELDQFCQGLTIGNVTHGLRSW IGWVVPAAATGGVATGALGLDCHFCFVPLHREAVTRPENAAASRAVLAKVGFREE GLRLRYLEVDRWRDLMLAITVEVYGVASVTLVRAGHASWP" 5975. .7150  
/gene="MT1025"  
5975. .7150  
/gene="MT1025"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45272.1"  
/db\_xref="GI:13880591"  
/translation="MNYRCVIALGALTFRPWPWTPMGPRDNRNRREQVINPISQSLIMI

```

SLVVLWLFVLVPLMISKRDVARTSDVALTRVLNGGAGARLLKRGGAACHRWGYLP
PFGQDDPKWKEEDNRDPVEDGFADVEHDDIODEQADRRARRKAVVKKVAAPTAG
ADEPFDYDVVVEEDSEALPVGAGAVGSADEADAADVAGHADPEADPEVEYE
YVEYDTCGLEEDDEQAPPTVAGSTSRREFDTKTAADVSAKYTKRKRALVYAV
IIVGSAAFELTPVAMWTCGATGVTVLYLRYLRQTRIEKVRRRMQRITARLQG
VENTRDREYDVVPSRLRRPCAVVLEIDDEDPITFTHLESAAPRTNYGWRPDLPRVAVGQ"
7201..7276
/gene="trNA-Ala-2"
7201..7276
/gene="trNA-Ala-2"
/product="trNA-Ala"
7324..7464
/gene="Mt1025.1"
7324..7464
/gene="Mt1025.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45273.1"
/db_xref="GI:13880592"
/translation="MVETASTALGAILAGNAVVPVPCPRRGAGLVVPSAVATAIWR
EA"
7550..7801
/gene="Mt1025.2"
7550..7801
/gene="Mt1025.2"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45274.1"
/db_xref="GI:13880593"
/translation="MSTKYLOKVPVEAVQPGFSLAIPHGDYRLFQVDCQMCQRSQ
OPWIMLMSSEVDGQPFVLEAGTAVIRLLGVCOAAS"
complement(7862..8071)
/gene="Mt1025.3"
complement(7862..8071)
/gene="Mt1025.3"
/notes="identified by Glimmer2; putative"
/codon_start=1

Query Match 100.0%; Score 25; DB 1; Length 20916;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCGGCGAGCCACCGAT 25
Db 6827 CCGCGCGGCGAGCCACCGAT 6803

RESULT 2
BX248337/c
LOCUS BX248337 327650 bp DNA linear BCT 11-JUN-2003
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
4/14.
ACCESSION BX248337 BX248333
VERSION BX248337.1 GI:31617663
KEYWORDS complete genome.
SOURCE Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

1
REFERENCE
AUTHORS
Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pyor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )

```

# REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 327650)  
Garnier,T.  
Direct Submission  
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire  
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex  
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham Lane, New Haw Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

## FEATURES source

1.327650  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
698..2170  
/locus\_tag="Mb0912"  
698..2170  
/locus\_tag="Mb0912"  
/note="Mb0912, -, len: 490 aa. Equivalent to Rv0888, len:  
490 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 490 aa overlap). Probable exported  
protein. Equivalent to AAK45157.1 from Mycobacterium  
tuberculosis strain CDC1551 (507 aa) but shorter 17 aa.  
Contains possible N-terminal signal sequence."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE EXPORTED PROTEIN"  
/protein\_id="CAD93773.1"  
/db\_xref="GI:3167664"  
/translation="MDYAKRIGVNGALAVLVGCAVTHAIGSAAPTDPSSSTDS  
VDACSPGGSSASSLAAIPGASVPGVQVDPGSIIPDILLNALIDFAAVANGVPII  
ENRTPVANPQSVFEGTVGVRDPCDPGNRMITFAVRERAPGPGQHGIVTDQR  
TASFIYADPGFVGTDTFSVNSDDTLRHVHGLAGYLGPPHGDVAVTVVFGNTPT  
DTISGDFSMLTNTIAGLPFLSSALPFLFFYTKIEIGRLNAYYVAVQEDFAYHQFLI  
KKSMPQTPPEPTLLMPGIVPFSGLNTLSEFKVQLRDTQWECTSDNCLTKGFI  
TYSQWRLPGGTVTVNLHTNGGPTTANLAQVANYIQNSAGRAVITVDFNARY  
SDDSALLOFAQVNLTDAMVQVPHGPTTPPTPCVMGNECELDDKIFYSRGGQVTL  
QAVSYGNEAPKPFNSKGEPLSDHSPAPVGVGHVADNAVVR"  
complement(2205..3326)  
/gene="cita"  
/locus\_tag="Mb0913c"  
complement(2205..3326)  
/gene="cita"  
/locus\_tag="Mb0913c"  
/EC\_number="2.3.3.1"  
/note="Mb0913c, cita, len: 373 aa. Equivalent to Rv0889c,  
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 373 aa overlap). Probable cita  
(alternate gene name: gita), citrate synthase 2 (EC  
4.1.3.7), highly similar to others e.g.  
CAB95899.1|AL359888 putative citrate synthase from  
Streptomyces coelicolor (387 aa); P39119|CISY BACSU  
citrate synthase II from Bacillus subtilis (366 aa), PASTA  
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa  
overlap); etc. Also similar to Rv0896|MTCY31.24 from  
Mycobacterium tuberculosis (29.2% identity in 274 aa  
overlap) and Rv1131. Contains PS00480 Citrate synthase  
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."  
/codon\_start=1  
/transl\_table=11  
/product="PROBABLE CITRATE SYNTHASE II CITA"  
/protein\_id="CAD93774.1"  
/db\_xref="GI:31617665"  
/translation="MTVVPEFVPLDGVAVFTTEIAPDKDGGALRYGVYIDELVLS  
ORVFGDWALLVGNFGSSGLPPAEPPLPIHSGDVVDVQAGLAMPALMGVYAPLLD  
IDDATAQQALARSVMALSYVAQSGLYQPAVFIQAVFIDECSVTAFRMFWQSEPDF  
RHIEAIDAYWVWSAAEHGNNASTFTARVIATGADVAAALSAGIANGSGPLHGGAPRV  
LPMLEVERAGDARSVVKGIIDRGEKLMGFGRHYRVRAEDPDPARVLRRAAERLGRPYE

gene

CDS

VAVVQAAALSELSEERRPRDAIETNVFWAAVVLDFARVPANMPPAMFTCGRTAGWCA  
HILEQRLGLKLVPSAIYVPGSPESVDGWERVLTTA"  
complement (3413..6061)  
/locus\_tag="Mb0914c"  
complement (3413..6061)  
/locus\_tag="Mb0914c"  
/notes="Mb0914c", len: 882 aa. Equivalent to Rv0890c,  
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.8% identity in 882 aa overlap). Probable  
transcriptional regulatory protein, LuxR family, highly  
similar (but shorter 238 aa in N-terminus) to  
NP\_302202.1|NC\_002677 possible transcriptional regulator  
from Mycobacterium leprae (1106 aa). Also highly similar  
(generally in part) to others e.g. T50568 probable  
multi-domain regulatory protein from Streptomyces  
coelicolor (1334 aa); P10957|NARL ECOLI nitrate/nitrite  
response regulator protein from Escherichia coli (216 aa),  
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99  
aa overlap); etc. Also highly similar to others from  
Mycobacterium tuberculosis e.g. MTCY02B10.22, MTW008.44,  
MTW036.21, and MTCY31.24. Contains PS00017 ATP/GTP-binding  
site motif A (P-loop), PS00622 Bacterial regulatory  
proteins, luxR family signature, and probable helix-turn  
helix motif from aa 836 to 857 (Score 1559, +4.50 SD).  
BELONGS TO THE LUXR/URPA FAMILY OF TRANSCRIPTIONAL  
REGULATORS."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN  
(PROBABLY LUXR-FAMILY)"  
/protein\_id="CAD93775.1"  
/db\_xref="GI:31617668"  
/translation="MALLAQNRVLTCGTGVKTRLAIQIASSELRDGLCFVFLA  
PITGESIVATRAVAPLDQPGESTWDSLEFIGNRMVLWLDNCEHLLDCAALVH  
ELIGACPEITLATSREPIAGCEITWRVPSMSTIDEAVELFADRASRVQCFITANH  
NAAVDSHCRLLDGIPLAFFAAARVSRSPLEIADLDDCFLLAGGVGAVQROQT  
LRASIDWHLITETQILFRLAPFVGGDFDLAAVRAVAGSDLPFSLDOLTLLVD  
KSLVADDCQGRVYLLRTPVRALEKIGDSGEADVHARHDYVTAASLNTPADN  
DHQRLVARATEIDNRNAPFASRENGHTEALQLASSLOPIWFGRAHLREGLSWFS  
ILEQPHRLATSVARALADKMTLWLTSPVGTDIIPAAQALAMAREVGGDP  
AALVRALATACGGSGYNAAPAFYFAEATDLARADIDRWLTCLQILYWRGVGTCISGDP  
NALRAAAEERCDLADITIGRFVSRHCSLMSLAQWAGNLTEALBSREITAEASRN  
DVTQKGLYTOAVLYACAGAAHAIAGACATAATGELGVYQIGIGYAMTAYALAAQ  
DVTAALEASDARPIRAQPDQVTHQVLMQALAGDPAIAROFANDAVDTANGWH  
RVALTIRAVATARGEPLARDADAALACGAELHIYQMPDEMLLAGELAGEVSH  
SEGVRLGAAALQRQTRVRFKIMDAGVQASVTLREAGWDEGDFDRAWEGAALSTH  
EATAYARQGERKRRPARGSLSTPTFTRDVRVRLVSEGLSKNDIAKELFVSPRTQTHL  
THYVAKGLASRVQLVDEAARRGSPS"  
complement (6063..6920)  
/locus\_tag="Mb0915c"  
complement (6063..6920)  
/locus\_tag="Mb0915c"  
/notes="Mb0915c", len: 285 aa. Equivalent to Rv0891c,  
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,  
(99.6% identity in 285 aa overlap). Possible  
transcriptional regulatory, highly similar in N-terminus to  
NP\_302202.1|NC\_002677 possible transcriptional regulator  
from Mycobacterium leprae (1106 aa). Also highly similar  
to several Mycobacterium tuberculosis putative  
transcriptional regulators e.g. Q1102|MTCY02B10.22  
PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),  
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in  
247 aa overlap); MTW036.21; MTW008.44; MTCY02B10.23. Also  
shows similarity with several adenylate cyclases and  
hydrolases from other organisms."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN"  
/protein\_id="CAD93776.1"  
/db\_xref="GI:31617667"  
/translation="MLFNVAHSLPPNIDHAILRGEDHPPTCAKVARGRISALGS  
LDLRYSLKCYAAPDPVGCSEFVPPRRVLIANOGLDVSRLEPTGTVTLLADVSEST  
HLWQPCPDMAIAHLDTVSEAITNHGQVQVRYEGDSFVAATRASDAACALD

LQRTSLAFLIRLIGLHTEGVLDELVYGTINRTARLRLDIAHGQVLSAATGDLVT  
GRLPADAWLVDLGRHPLRGLPRPWWQLCHPDIREKFPPLRTAKSSPTSLPAQFTT  
FVGERAOIS"  
7318..8805  
/locus\_tag="Mb0916"  
7318..8805  
/locus\_tag="Mb0916"  
/EC\_number="1.14.-.-"  
/note="Mb0916", len: 495 aa. Equivalent to Rv0892, len:  
495 aa, from Mycobacterium tuberculosis strain H37Rv,  
(100.0% identity in 495 aa overlap). Probable  
monooxygenase (EC 1.14.-.-), highly similar to others e.g.  
NP\_250787.1|NC\_002516 probable flavin-binding  
monooxygenase from Pseudomonas aeruginosa (491 aa);  
monooxygenase from Streptomyces  
CAP59668.1|AL132674 monooxygenase from Streptomyces  
coelicolor (519 aa); P12015|CYMO ACIS cyclohexanone  
monooxygenase from Acinetobacter sp. (542 aa), FASTA  
scores: opt: 489, E(): 6.8e-25, (30.3% identity in 492 aa  
overlap); etc. Also highly similar to Rv0585c, Rv3854c,  
Rv3083, etc from Mycobacterium tuberculosis. Has  
hydrophobic stretch at N-terminus."  
/codon\_start=1  
/transl\_table=1  
/product="PROBABLE MONOOXYGENASE"  
/protein\_id="CAD93777.1"  
/db\_xref="GI:31617688"  
/translation="MTGRCPTVAVVAGMSGMCVAILLSAGITDVCIVYEKADVGGT  
WRDNTYPLGTCDFPSLYQYSPAKFNWTFMRGGEIODYLGRIGRGLRIRIFG  
ADTVSARFDDGRWLRDTSGETVLDPLISATGVLHPRIPIIAGLDDFRGTVFHSAR  
WDHTVPLGRRIAVIGTSGTVGLCVGLAGVAGVMTFORTQWLVLPWNPRTYSLKAR  
VFHEAFPLCLGLAYKAYSLSEFVALSNGLHRLKLVAGVACRSLRRVREDPRLRAL  
TPDEPFCNKLVMGSGFYRAIQRDDVELVTAGIDHVEHRCIVITDDGVLEHVDVILAT  
GFDSHAFRPNQUTRGDIRIDDDWQDPHATQVLPFPNFFMGLPSPVGNFPL  
TAVAESQAELHVQWIKWRHGEFTDEPKSAATEAYNTVLRAAMPNTVVTGCDSWYL

Query Match 100.0%; Score 25; DB 1; Length 327650;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGACCCACCGAT 25  
Db 126588 CCGCGCGCGCAGACCCACCGAT 126564

RESULT 3  
BX842575/c standard; circular genomic DNA; PRO; 349306 BP.  
ID BX842575 AC BX842575 AL010186; AL021006; AL021897; AL023456; 279700; 292539;  
AC BX842575; AL010186; AL021006; AL021897; AL023456; 279700; 292539;  
AC Z93777; Z94752; Z95209; Z95210; Z95584; Z95585; Z98260;  
XX  
SV BX842575.1  
XX  
DT 21-NOV-2003 (Rel. 77, Created)  
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)  
XX  
DE Mycobacterium tuberculosis H37Rv complete genome; segment 4/13  
XX complete genome.  
XX  
OS Mycobacterium tuberculosis H37Rv  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.  
XX  
[1]  
RX MEDLINE; 98295987.  
RX PUBMED; 9634230.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmier K., Gas S., Barry III C.E., Tekala F., Badcock K.,  
RA Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,  
RA Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,  
RA Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,



```

FT misc_feature 2844..2930
FT /notes="PS00402 Binding-protein-dependent transport systems
FT inner membrane comp signature"
FT CDS
FT 3212..4138
FT /notes="Rv0930, (MTCY21C12.24), len: 308 aa. Probable pstA1,
FT phosphate-transport integral membrane ABC transporter (see
FT citation below), highly similar to others e.g.
FT NP_302393.1|NC_002677 membrane-bound component of phosphate
FT transport from Mycobacterium leprae (304 aa);
FT CAB88473.1|AL353816 phosphate ABC transport system permease
FT protein from Streptomyces coelicolor (354 aa) (N-terminus
FT longer); NP_312689.1|NC_002695 phosphate transport system
FT permease protein PstA from Escherichia coli strain O157:H7
FT (296 aa), etc. Also similar to Rv0936|MTCY0809.03c|PSTA2
FT PROBABLE TRANSMEMBRANE ABC TRANSPORTER COMPONENT OF
FT PHOSPHATE UPTAKE SYSTEM from Mycobacterium tuberculosis
FT (301 aa)."
FT /transl_table=11
FT /function="INVOLVED IN ACTIVE TRANSPORT OF INORGANIC
FT PHOSPHATE ACROSS THE MEMBRANE (IMPORT); RESPONSIBLE FOR THE
FT TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE. THIS IS
FT ONE OF THE PROTEINS REQUIRED FOR BINDING-PROTEIN-MEDIATED
FT PHOSPHATE TRANSPORT."
FT /genes="pstA1"
FT /locus_tag="Rv0930"
FT /products="PROBABLE PHOSPHATE-TRANSPORT INTEGRAL MEMBRANE
FT ABC TRANSPORTER PSTA1"
FT
FT Query Match 100.0%; Score 25; DB 15; Length 349306;
FT Best Local Similarity 100.0%; Pred. No. 3.3;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGGCGGAGAGCCACACGAGT 25
Db 79350 CCGCGGCGGAGAGCCACACGAGT 79326

RESULT 4
MLC1373/c 37304 bp DNA linear BCT 27-AUG-1999
LOCUS
Myobacterium leprae cosmid L373.
ACCESSION
AL033500
VERSION
AL033500.1 GI:4455688
KEYWORDS
acyl-CoA carboxylase; acyl-CoA dehydrogenase; acyltransferase;
ATP-dependent DNA helicase; biotin carboxylase; enoyl-CoA
hydratase; galU; gated mechanosensitive ion channel; membrane
protein; moeA; molybdenum cofactor;
phosphoribosylaminoimidazolecarboxamide formyltransferase;
phosphoribosylglycinamide formyltransferase; protein kinase;
pseudogene; purH; RLEP; serine protease; succ; succinyl-CoA
synthetase alpha subunit; succinyl-CoA synthetase beta subunit;
sucD; two-component response regulator; two-component system
sensor; UTP-glucose-1-phosphate uridylyltransferase.
SOURCE
Mycobacterium leprae
ORGANISM
Mycobacterium leprae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE
1 (bases 1 to 37304)
Eiglmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
JOURNAL
93188700
MEDLINE
8446027
PUBMED
2 (bases 1 to 37304)
REFERENCE
Harris, D. and Taylor, K.
Unpublished
JOURNAL
3 (bases 1 to 37304)
REFERENCE
James, K.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
JOURNAL
Submitted (17-FEB-1998) Mycobacterium leprae sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.

```

Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France Requests for cosmids should be sent to Karin Eiglmeyer (kei@pasteur.fr)

Notes:  
The Sanger Centre is funded to complete the sequence of M. leprae by the Heiser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.  
Work in Paris is supported by the Heiser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIP-GREG).  
Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, <http://www.sanger.ac.uk/Projects/>)  
CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), CB33 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

## FEATURES

```

Location/Qualifiers
1..37304
1..37304
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/db_xref="taxon:1763"
/db_xref="cosmid L373"
1..755
/gene="MLCB373.01"
<1..755
/gene="MLCB373.01"
/notes="MLCB373.01, probable ATP-dependent DNA helicase, partial CDS, len: >250 aa; similar to many e.g. PCRA_BACST, Bacillus Stearothermophilus DNA helicase pcra (724 aa), fasta scores; Opt: 402 z-score: 476.0 E:1 3.3e-19, 40.3% identity in 248 aa overlap and UVRD_ECOLI (EMBL:N87049) E.coli DNA helicase II (720 aa) (39.8% identity in 196 aa overlap). Equivalent to TR:P71561 (EMBL:Z79700) Rv0949 (MTCY10D7.25c) M.tuberculosis probable DNA helicase (771 aa) (80.2% identity in 248 aa overlap)"
/codon_start=3
/transl_table=11
/product="probable ATP-dependent DNA helicase"
/protein_id="CAB36663.1"
/db_xref="GI:4455689"
/db_xref="GOA:Q925T0"
/db_xref="SPREMBL:Q925I0"
/translation="PQELARDLNINELVFAHEFSTEQANAALAKSLHTPEDEYDPTGALAFLEKVSLSMDTDQIPENNSGVITLTHAAKLEFFVYFVTGWEDGKLPKMTGLDPTLSERLAVYTRARQRLYLRATRSWGQPIINPESRFLREIPPELIDWRSILTDYSYTPASGASRGRVPSIRSGASKRALLVLAPGDRVTHDKYGLGRVEEVSGVGSATSLIDFGSSGRKLMHNHAPVKL"
/complement (1307..2338)
/gene="MLCB373.02c"
/complement (1307..2338)
/gene="MLCB373.02c"
/notes="MLCB373.02c, hypothetical protein, len: 343 aa;

```



unknown function, similar to regions of many bacterial hypothetical proteins e.g. to the C-terminus of TR:P74517 (EMBL:P90915) from *Synechocystis* sp. strain PCC6803 (715aa), fasta scores; opt: 446 z-score: 499.0 E(1): 1.8e-20, 52.5% identity in 122 aa overlap. Equivalent to TR:P71560 (EMBL:P279700) Rv0950c (MTCY10D7.24) M.tuberculosis hypothetical protein (332 aa) (71.0% identity in 331 aa overlap)" /codon\_start=1 /transl\_table=1 /product="hypothetical protein MLCB373.02c" /protein\_id="CAB36664.1" /db\_xref="GI:4455690" /db\_xref="GI:4455690" /db\_xref="SPTREML:Q9ZSH9" /translation="MPQHDILVRFCAAEAVRSFRRNLNQRNDVTETIPVDGDDDFD LVDYDLFAEDIEQFSDDDYFCNETQVLLASEIDLDQDLSLASELAASANVAFNP TSAPRRPCQKQKQVTSRGLLIISAMAAGAMATAHTVISHDDIIPKTEITLVANASA LSGATVQDQGVQVITPAASTAVNEEHAKTQFAFAHQAQREARLSLOPLVMPKTK GIFTNFTGYSYGVNLHAGIDLANAIGTPIILAVSDGVVIDGPTAGYGMWKLKHADCTV TLGYNSFTLVNNGQYVVGADQIATMTGRGNSGPHLHFEVLSSGERIDPVPWLAKR GLIVGNTG" /gene="sucC" /note="MLCB373.03, sucC, probable succinyl-CoA synthetase beta subunit, len: 393 aa; similar to many bacterial succinyl-CoA synthetases e.g. SUCC\_TREFL (EMBL:X54073) from *Thermus aquaticus* B (378 aa), fasta scores; opt: 810 z-score: 894.5 E(1): 0, 48.7% identity in 384 aa overlap. Equivalent to SUCC\_MYCTU (EMBL:P279700) Rv0951 (MTCY10D7.23c) M.tuberculosis sucC (387 aa) (86.7% identity in 391 aa overlap) Contains Pfam match to entry PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35 and PS01217 ATP-citrate lyase / succinyl-CoA ligases family signature" /codon\_start=1 /transl\_table=1 /product="probable succinyl-CoA synthetase beta subunit" /protein\_id="CAB36665.1" /db\_xref="GI:4455691" /db\_xref="GI:4455691" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KAQVQGRGKAGKVAATPEEAYQAKNIIIGDIDKGIHVKKLLVVEASDIAEYPL SFLLDRANRTVLAMCSVEGMEIEEVAATPERIAKVSVDVAGVDLACARSIAQOQH LPASVLDAAATIAKIELFVAEDATVEVNPILVTPSRGCVSGSKTLADLGGKVTIDA NARTPQPHAEFFEDRAITPLEWAKAKOHNLNLIKLDGEVGIIGNAGLTMTSEVDVAY AGEQHGKVPANFLDIIGGSAEAYNAASLDVLGDQGVKSFVFNIFGGITICTDTATG IVKALEILGDEANKPLVVRLDGNNVGEGRILAEANPLVMLVPTMDEAAANKAAELAR T" /gene="sucC" /note="PF00549 ligase-CoA, CoA-ligases, score 129.00, E-value 6.9e-35" /db\_xref="GOA:Q9ZSH8" /db\_xref="SWISS-PROT:Q9ZSH8" /translation="MDLFYQAKELFKHGVPGIQGRVTSDEGAKTIAMEIRPVMV KA

ATP-citrate lyase / succinyl-CoA ligases family signature,  
PS00399 ATP-citrate lyase / succinyl-CoA ligases family  
active site and PS00017 ATP/GTP-binding site motif A  
(P-loop)"  
/codon\_start=1  
/transl\_table=11  
/product="probable succinyl-CoA synthetase alpha subunit"  
/protein\_id="CAB36666.1"  
/db\_xref="GI:4455692"  
/db\_xref="GOA:O9S372"  
/translation="NAIFLTKONTNVIQGITGSEAAVHARMKAGTQIVGGVNAARKA  
GVNTHGPKDQSIITLPVFGTVAEATIGNCPGIIIPVPRPAKEAIIBAVEAEIPL  
VIITGPIVQDSVYAWYNLNGTKTRIIGNCPGIIIPVPRPAKEAIIBAVEAEIPL  
LVSKSTLYIQMFELSDRGSTAGIGSDPVIGTHIDAIENAEPEODPTKIIIVMIG  
IGADVEAADYIKANSPVGVIVAGTFAPEGTGTHAGAIIVSGSGGTAIAVKKDALE  
IAGVKGKTPSETAVLVRILCTL"  
misc\_feature  
4235..4672  
/gene="sucD"  
/note="Pfam match to entry PF00549 ligase-CoA,

Query Match 87.2%; Score 21.8; DB 1; Length 37304;  
Best Local Similarity 92.0%; Pred. NO. 85;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGAGCCACCAAGAT 25  
|||||  
DB 35073 CCGCTGGCGGAGAGCCACCAAGAT 35049  
|||||

RESULT 5  
MLEPRTN1/c  
LOCUS  
DEFINITION Mycobacterium leprae strain TN complete genome, segment 1/10.  
ACCESSION AL583917 AL450380  
VERSION AL583917.1 GI:13092412  
KEYWORDS  
SOURCE  
ORGANISM Mycobacterium leprae  
Mycobacterium leprae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
1 (bases 1 to 344050)  
Cole, S.R., Eigmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,  
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,  
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,  
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,  
Hamlin, N., Holtroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,  
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,  
Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,  
Simmonds, M., Skellon, J., Squares, R., Squares, S., Stevens, K.,  
Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G.  
Massive gene decay in the leprosy bacillus  
Nature 409 (6823), 1007-1011 (2001)  
21128732  
PUBMED  
11234002  
2 (bases 1 to 344050)  
Parkhill, J.  
Direct Submission  
Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium  
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome  
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75722, Paris Cedex, France. E-mail: parkhillsanger.ac.uk  
Notes:  
Details of M. leprae sequencing at the Sanger Centre are available  
from [http://www.sanger.ac.uk/projects/M\\_leprae/](http://www.sanger.ac.uk/projects/M_leprae/) A relational  
database containing the M. leprae sequences is available from  
<http://genolist.pasteur.fr/leprae/>.  
Location/Qualifiers  
1..344050  
/organism="Mycobacterium leprae"  
/mol\_type="genomic DNA"  
/strain="TN"

FEATURES  
source

```

/db xref="taxon:1769"
1..1566
/gene="dnaA"
/notes="synonym: ML0001"
1..1566
/gene="dnaA"
/notes="Similar to M. tuberculosis dnaA, chromosomal
replication initiator protein, SW:DNAA_MYCTU (P45993) (507
aa); Fasta score E(): 0, 87.2% identity in 507 aa overlap
and Mycobacterium smegmatis dnaA, SW:DNAA_MYCSM (P49992)
(495 aa); Fasta score E(): 0, 69.2% identity in 441 aa
overlap. Previously sequenced as SM:DNAA_MYCLE (P46388)
(521 aa); Fasta score E(): 0, 99.8% identity in 521 aa
overlap. Contains Pfam match to entry PF00308 bac dnaA,
Bacterial dnaA protein. Contains PS00017 ATP/GTP-binding
site motif A (P-loop). Contains PS01008 DnaA protein
signature."
/codon_start=1
/transl_table=11
/product="putative chromosomal replication initiator
protein"
/protein_id="CAC29509.1"
/db xref="GI:13092413"
/db xref="GOA:P46388"
/db xref="SWISS-PROT:P46388"
/translations="MFVPHAKKPEIYENQDTSGLADLSIGFTTVMNAVVSINQESN
TDBEATNDSTLVTLPQORALNLVQPLTIEGALLSVPSFVQNEIERLRTPI
DALSRLLGOIOIGVRIAPSTDHIDNDSADVLTDGCTDENVGEPITGEVQ
LPYIFERPHETSTVTGTSNLNRYTFEIVGASNRFAHAALAAIAAPARAPNL
FIWESGLGTHLLHAAGNAQVFPGMKVYVSTEEFTNFDNSLRDVRKVFKRSY
RDVLLVDIOPIEGKEIGIEFTFTLNANKQIVISSDRPPQIATLEDRLT
RFEWGLTDVQPELLETRIALKQMERLAVPGDVLIELIASIERNIIELEGALIR
VTPASLNKTAIVLADIASTWQISATIMTAAYFDITIEELRGPK
TRALASQROIAMVLCRELTDLSLPKIQAFGRDHTTVMYQAKILSEAREVEFDVH
KELTTRIRSKR"
544..1482
/gene="dnaA"
/notes="Pfam match to entry PF00308 bac dnaA, Bacterial
dnaA protein, score 712.20, E-value 8.3e-240"
664..687
/gene="dnaA"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
1426..1482
/gene="dnaA"
/notes="PS01008 DnaA protein signature"
2081..3280
/gene="dnaA"
/notes="synonym: ML0002"
2081..3280
/gene="dnaA"
/ECnumber="2.7.7.7"
/notes="Similar to M. tuberculosis dnaN, DNA polymerase
III, [beta] subunit, SW:DP3B_MYCTU (Q50790) (402 aa);
Fasta score E(): 0, 80.9% identity in 403 aa overlap and
Mycobacterium smegmatis dnaN, SW:DP3B_MYCSM (P52851) (397
aa); Fasta score E(): 0, 77.3% identity in 397 aa overlap.
Previously sequenced as SW:DP3B_MYCLE (P46387) (399 aa);
Fasta score E(): 0, 100.0% identity in 399 aa overlap.
Contains Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit."
/codon_start=1
/transl_table=11
/product="putative DNA polymerase III, [beta] subunit"
/protein_id="CAC29510.1"
/db xref="GI:13092414"
/db xref="GOA:P46387"
/db xref="SWISS-PROT:P46387"
/translations="MDLATNVCGLKFLAREGFASVSWVAKYLPTRTPVPVLGS
VLLTGSGLTSGFDYEVSAEVOAAEIASGSLVLSGRLLSDITRALPNKPVHFVY
DGNVALTCGSARFSLPTWAVEDYPLTLPDCTGLPSDFEAFAGOVATAAGSDYT
LPMGLTIRIEISGSDTVVLAATRFLANRELKXSVLSDPFEASVLVPKTLVEYAKG
TDGSGVCLSGVGVGKGLFGISGGGRKSTRLLDAEFFPKFROLLEPAENTAVATID
VAELTBAIKLVALVADRGAQVRMEFGDGLIRLSAGADDDVGRABEDLAVFTGEPLTIA
ENPNYLTDLGLASVHSERSVFGFTTFSKPEALLRPTSDNDVHPTDGPFPALPTDYVILL
MPVRLDG"
2117..3214
/gene="dnaN"
/notes="Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit, score 592.70, E-value
2.2e-174"
2..174"
3281..4438
/gene="recF"
/notes="synonym: ML0003"
3281..4438
/gene="recF"
/notes="Similar to M. tuberculosis recF, DNA replication
and SOS induction protein, SW:RECF_MYCTU (Q59586) (385
aa); Fasta score E(): 0, 76.4% identity in 385 aa overlap
(384 aa); Fasta score E(): 0, 70.0% identity in 383 aa
overlap. Previously sequenced as SW:RECF_MYCLE (P46391)
(385 aa); Fasta score E(): 0, 99.5% identity in 385 aa
overlap. Contains 2 Pfam matches to entry PF00470 RecF,
RecF protein. Contains PS00617 RecF protein signature 1.
Contains PS00618 RecF protein signature 2."
/codon_start=1
/transl_table=11
/product="putative DNA replication and SOS induction
protein"
/protein_id="CAC29511.1"
/db xref="GI:13092415"
/db xref="GOA:P46391"
/db xref="SWISS-PROT:P46391"
/translations="MYVHFGLRDFRSDHVDLNLPGRTVFRFGNGKTNLEIALM
RSTLSSHRYGTDTPLIRAGTIRAVSTIVNREGCAITDEIAAGANRANRSLIV
RGMEVGVRLARVLFAPEALVAGDPANRRYVLDLQATVROPVIAVRADYDKVLQ
RTALKSLAARVRSQGVLDLTDVMDTLEAHEGALMAARIQVNLQALPEVKVLQ
LAPGSRASISYRASLDIGIAGVSGSDRALQDLGLSTRNVELEGICLVGPH
RDELELRGQPKAFASGHGWSLAIALAAYELLRAGNDFVLLDDVFAELDNA
RCRALATVAESAEOQLVTSAAQEDIPVGWDAKWTVDLRDSDSGRVSVVYP"
3287..3415
/gene="recF"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 18.90, E-value 0.00024"
3368..3391
/gene="recF"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
3614..3691
/gene="recF"
/notes="PS00617 RecF protein signature 1"
4100..4423
/gene="recF"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 5.20, E-value 1.5"
4244..4297
/gene="recF"
/notes="PS00618 RecF protein signature 2"
4435..5004
/gene="ML0004"
/notes="ML0004"
/notes="ML0004"
/notes="Similar to hypothetical proteins from mycobacteria
e.g. M. tuberculosis RV0004, hypothetical protein,
TR:P71573 (EMBL:AL123456) (187 aa); Fasta score E(): 0,
77.2% identity in 167 aa overlap. Previously sequenced as
TR:Q50181 (EMBL:Z70722) (199 aa); Fasta score E(): 0,
99.5% identity in 187 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC29512.1"
/db xref="GI:13092416"
/db xref="SWISS-PROT:Q5CDF4"
/translations="MIESNESISYGGDTTPELTGLSGFDLVRRLAEARAAACQKQD
AGRHVVVPPVFPVFRVRRNWSGFPDVPDQPLGKVAHDLAKRGWSAQVGRVFG

```

```
OWASVVGGOIADHAPPVGLNGLSVTAESTAWATOLRIMOQALLAKIAAAVGVNGVWT
SLKITCTPTAPSRKRGFWHTAGRPRTYTG"
5211..5216
/notes="possible RBS"
gene 5229..7265
/genes="gyrB"
/notes="synonym: ML0005"
CDS 5229..7265
/genes="gyrB"
/EC_number="5.99.1.3"
/notes="Similar to M. tuberculosis gyrB, DNA gyrase subunit
B, SW:GYRB_MYCTU (P41514) (686 aa); Fasta score E(): 0,
87.5% identity in 679 aa overlap and to Mycobacterium

Query Match 87.2%; Score 21.8; DB 1; Length 344050;
Best Local Similarity 92.0%; Pred. No. 78; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGAGCCACACAGGAT 25
Db 250926 CCGCTCGCGCAGAGCCACACAGAT 250902

RESULT 6
AF063800/c
LOCUS AF063800
DEFINITION Odontella sinensis glyceraldehyde-3-phosphate dehydrogenase
precursor (GapCl) mRNA, complete cds.
ACCESSION AF063800
VERSION 20142334
KEYWORDS Odontella sinensis
SOURCE Odontella sinensis
ORGANISM Odontella sinensis
REFERENCE 1 (bases 1 to 1309)
AUTHORS Liand,M.F., Lichtle,C., Apt,K., Martin,W. and Cerff,R.
TITLE Compartment-specific isoforms of TPI and GAPDH are imported into
diatom mitochondria as a fusion protein: evidence in favor of a
mitochondrial origin of the eukaryotic glycolytic pathway
Mol. Biol. Evol. 17 (2), 213-223 (2000)
JOURNAL 20142334
MEDLINE 10677844
PUBMED 10677844
REFERENCE 2 (bases 1 to 1309)
AUTHORS Liand,M.F.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-1998) Institut fuer Genetik, Technische
Universitaet Braunschweig, Spielmannstrasse 7, Postfach 3329,
Braunschweig D-38106, Germany
FEATURES
source Location/Qualifiers
1..1309
/organism="Odontella sinensis"
/mol_type="mRNA"
/db_xref="taxon:2839"
gene 1..1309
/genes="GapCl"
5'UTR 1..7
/genes="GapCl"
CDS 8..1138
/genes="GapCl"
/notes="GAPDH"
/codon_start=1
/product="glyceraldehyde-3-phosphate dehydrogenase
precursor"
/protein_id="AAP34326.1"
/db_xref="GI:6979052"
/transl_table="MKSIAALIVGSAAYAFPSFSGTSLKASSNGASMSMTGKGVN
GFGRIKLTIRIMWEDCSLAINAGSATPDYMAVQYKDTIHGIAGSKVIYDGFL
VLNGKKIQTSCREDPKEVGMGALGADYVCSTGVFTKEKAQAIIDGGAKVIYGA
KDSQIVGVNQAEYDGEDFISCSCTNGLA PMVKAIHDFEVLBEALMTVHMT
ATQAVDSSCKDWRGGRASGNIIIPSSGAKAVTKYIPSLQGLKGVAFVPIDV
SVVDLTLEKATTEIICAYVKSAGDMKGLGYSDEPLVSTYEGNTTSSIFDAD
AGIMLPNPFVKLVAVYDWNWGVSCRVVDLMKHAVAADKAVEA"
```

```
transit_peptide 8..118
/genes="GapCl"
mat_peptide 119..1135
/genes="GapCl"
/product="glyceraldehyde-3-phosphate dehydrogenase"
1139..1309
/genes="GapCl"

3'UTR
ORIGIN
Query Match 81.6%; Score 20.4; DB 8; Length 1309;
Best Local Similarity 95.5%; Pred. No. 3.9e+02; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCGCGCGCAGAGCCACACAGGA 24
Db 51 CCGCGCGCAGAGCCACACAGGA 30

RESULT 7
HSA333853 732 bp DNA linear PRI 18-JUL-2002
LOCUS HSA333853
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
HST-EG10RS.
ACCESSION AJ333853
VERSION AJ333853.1 GI:15878271
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 732)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 732)
JOURNAL Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podewski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashuba,V.I.,
Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
NOTI flanking sequences: a tool for gene discovery and verification
of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
JOURNAL 22131767
MEDLINE 12136098
PUBMED 12136098
REFERENCE 2 (bases 1 to 732)
AUTHORS Zabarovsky,E.R.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
source Location/Qualifiers
1..732
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="HST-EG10RS"

ORIGIN
Query Match 80.8%; Score 20.2; DB 9; Length 732;
Best Local Similarity 88.0%; Pred. No. 4.8e+02; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGCGCGCAGAGCCACACAGGAT 25
Db 244 CCGCGCGCAGAGCCACACAGGAT 268

RESULT 8
BX640443/c
LOCUS BX640443
DEFINITION Bordetella bronchiseptica strain RB50, complete genome; segment
7/16.
ACCESSION BX640443
VERSION BX640443.1 GI:33575370
KEYWORDS complete genome.
SOURCE Bordetella bronchiseptica RB50
ORGANISM Bordetella bronchiseptica RB50
```

Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Alcaligenaceae; Bordetella.

# REFERENCE

1 Parkhill, J., Sebatia, M., Preston, A., Murphy, L.D., Thomson, N.,  
Harris, D.E., Holden, M.T.G., Churcher, C.R., Bentley, S.D.,  
Mungall, K.L., Cerdano-Tarraga, A.M., Temple, L., James, K., Harris, B.,  
Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N.,  
Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P.,  
Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H.,  
Holt, S., Jags, K., Leather, S., Moulé, S., Norbertczak, H.,  
O'Neill, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S.,  
Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M.,  
Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L.,  
Whitehead, S., Barrall, B.G. and Maskell, D.J.  
Comparative analysis of the genome sequences of Bordetella  
pertussis, Bordetella parapertussis and Bordetella bronchiseptica  
Online Publication  
Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227  
2 (bases 1 to 346274)

# FEATURES

1. 346274  
Location/Qualifiers  
/organism="Bordetella bronchiseptica RB50"  
/mol\_type="genomic DNA"  
/strain="RB50"  
/db\_xref="taxon:257310"  
100..918  
/locus\_tag="BB1951"  
100..918  
/locus\_tag="BB1951"  
/notes="ortholog of Bordetella pertussis (BX470248) BP2892"  
/codon\_start=1  
/transl\_table=11  
/product="putative integral membrane protein"  
/protein\_id="CAE32448.1"  
/db\_xref="GI:33575371"  
/translation="MDVWLICLLALGAVGVGAAGLLGGGMLLPFLTLMFAWQGM  
PGLVVAATATSMITSLTSSVRAHOKGTIOWSIWMAAPGIIIGLLSGGAVF  
AALSTAWLSLFFALFVGSWMMLRSKPKASQKQVGTSAAGAGIGLSIGVAG  
GGSLVFFWVCWVALHTAVSLSAALGFPIALANSVGVISLNSVTRPQMLGIYV  
PALALVATSLTAPLIGARMARLPVGTLRVFTLLFALAAINLTKAWQAFMV"  
100..156  
/locus\_tag="BB1951"  
/notes="Signal peptide predicted for BB1951 by SignalP 2.0  
HMM (signal peptide probability 0.995) with cleavage site  
probability 0.683 between residues 19 and 20;  
signal-peptide site"  
join(118..186,244..303,337..405,418..477,535..603,  
631..699,736..804,847..900)  
/locus\_tag="BB1951"  
/notes="8 probable transmembrane helices predicted for  
BB1951 by TMHMM2.0 at aa 7-29, 49-68, 80-102, 107-126,  
146-168, 178-200, 213-235 and 250-267"  
373..888  
/locus\_tag="BB1951"  
/notes="HMMpfam hit to PF01925, DE Domain of unknown  
function DUF81, score 4.1e-14"  
complement(1016..1627)  
/locus\_tag="BB1951"  
complement(1016..1627)  
/locus\_tag="BB1951"  
/notes="ortholog of Bordetella pertussis (BX470248) BP2891"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAE32449.1"  
/db\_xref="GI:33575372"  
/translation="MKLIGLSLSPYVKRVVVMASKLDDYOLENVHSPSQVOKYN  
PLGKVPCLVMDGGALFDSRVIVYEDTLSPVARLIPQGRERAIVKWEALDGLLD

# misc\_feature

ACVTWRLRPAQAQSPWTEROYGVKVRASLDAMDHSLGENAHCMGWNYSADIAGV  
CALCYLDRPADLMDRADHANLARLYEKLARSFVDSMPPTA"  
complement(1049..1387)  
/locus\_tag="BB1952"  
/note="HMMpfam hit to PF00043, DE Glutathione  
S-transferase, C-terminal domain, score 3.8e-07"  
complement(1412..1627)  
/locus\_tag="BB1952"  
/note="HMMpfam hit to PF02798, DE Glutathione  
S-transferase, N-terminal domain, score 6.1e-10"  
1742..3163  
/gene="purB"  
/locus\_tag="BB1953"  
1742..3163  
/gene="purB"  
/locus\_tag="BB1953"  
/EC\_number="4.3.2.2"  
/note="ortholog of Bordetella pertussis (BX470248) BP2890"  
/codon\_start=1  
/transl\_table=11  
/product="adenylosuccinate lyase"  
/protein\_id="CAE32450.1"  
/db\_xref="GI:33575373"  
/translation="MVKSPGLQPLPSTYDMQIADQLTQNLASPLDGRYASRAGALRG  
LSEAGFMAHREVEVAVLWALSDAGLPELPAPFAARERLALVRNFSADARND  
IERTVNDKVAVEYMLKEQVADDAELARAAEFTHFACTSDINTSHALMLTRARAV  
VPLQNLQAKLNEARTHADQPMLSRTHQGPASPTTLGKEFANHARLQRAVAATEA  
VEPLKNGATGNYNAHLSAYPEIDWPAFNRNVLAGLGTQNRHTQTIEPHDWAAALF  
DAVARNVILDLDRDINGVYALGVKRLKEGVSSVTPHKVNPDTFNSGNLGL  
ANAVRLHLSKLPVSRMORDLTSTVLRNLGVGLVALVADWACLGLNKLEVNAAL  
DADIDACWEVLAEPVQVMRYGLPQPEYEQKALTRGKITTEALREFIGGLALPDEP  
KARLLAMTPRSYIGLAAQLARAV"  
1838..3085  
/gene="purB"  
/locus\_tag="BB1953"  
/note="HMMpfam hit to PF00206, DE Lyase, score 2.1e-78"  
2675..2704  
/gene="purB"  
/locus\_tag="BB1953"  
/note="PS00163 Fumarate lyases signature."  
3226..3684  
/locus\_tag="BB1954"  
3226..3684  
/locus\_tag="BB1954"  
/note="ortholog of Bordetella pertussis (BX470248) BP2889"  
/codon\_start=1  
/transl\_table=11  
/product="putative c'cytochrome"  
/protein\_id="CAE32451.1"  
/db\_xref="GI:33575374"  
/translation="MKKLSIVAAVCTMLAPLWSSPAAAFKPODAIKYQSALTIM  
ASHFGKQPVNRGQAPYDAQIKRANVDIFKTLTLEWAAFGPTGEGDALPAVWDAE  
DFKQKQRLADNVAKUSAAADAGDYDKVRAAFGVGKACKACHDSFRKK"  
3226..3300  
/locus\_tag="BB1954"  
/note="Signal peptide predicted for BB1954 by SignalP 2.0  
HMM (signal peptide probability 1.000) with cleavage site  
probability 0.905 between residues 25 and 26;  
signal-peptide site"  
3316..3678  
/locus\_tag="BB1954"  
/note="HMMpfam hit to PF01322, DE Cytochrome C', score  
2.9e-56"  
3556..3606  
/locus\_tag="BB1954"  
/note="PS00238 Visual pigments (opsins) retinal binding  
site."  
3646..3663  
/locus\_tag="BB1954"  
/note="PS00190 Cytochrome c family heme-binding site  
signature."  
complement(3769..4464)  
/locus\_tag="BB1955"

# misc\_feature

# gene

# CDS

# misc\_feature

# misc\_feature

# gene

# CDS

# misc\_feature

# misc\_feature

# misc\_feature

# misc\_feature

# gene

```
CDS
complement(3769..4464)
/locus_tag="BB1955"
/notes="ortholog of Bordetella pertussis (BX470248) BP2888"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAE32452.1"
/db_xref="GI:33575375"
/translation="MQPTATSVRIWDLPTPLFHFNAFAAAVIGALVTVKLGGLYMDWHV
RGLTALGVLFRLIWFVGRPHAFATVFRGPGAMLYLRGAAADHAGHNPGLGALS
LMLAVIGVQAATGIFASDDIMQGLYGHVSEATANTLTGLQHANWIIIFGLGAHL
LAAVYSAVERRALVRAMITGNAPAAQLPPATPSAQDGFALWLRRAVVGACCAALVLM
IQAAVAVMDSPS"
misc_feature
complement(3799..3867,3958..4017,4108..4176,
4270..4338,4348..4416)
/locus_tag="BB1955"
/notes="5 probable transmembrane helices predicted for
BB1955 by TMHMM2.0 at aa 17-39, 43-65, 97-119, 150-169 and
200-222"
gene
complement(4526..4807)
/locus_tag="BB1956"
CDS
complement(4526..4807)
/locus_tag="BB1956"
/notes="ortholog of Bordetella pertussis (BX470248) BP2887"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE32453.1"
/db_xref="GI:33575376"
/translation="MNQLFDTNHLIRRPAGTKLDDTPIHVGQVHVLQGLADGPFVIRARV
IYNAPINGTITVTTELIRCAQDDKASHGQRINFRHEHVHIESVRLGTH"
5319..6122
/locus_tag="BB1957"
5319..6122
/locus_tag="BB1957"
/notes="ortholog of Bordetella pertussis (BX470248) BP2886"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAE32454.1"
/db_xref="GI:33575377"
/translation="MAIKLQGSFPDASPIMPMSGAANVLEGRFMRFGWGLPALA
VVLAAAGCVNREPEERAEIAYLEQVAPQAGVVAAPPDPTRKALGDEAYEYEMEA
AHAAVREALAAQALRLHSDIETVARQCDRLRAERLAAARTGLQEAAADAA
RAGMEQPPDLNAYARAYEHSVTAQALARI,SGLEPAVEDARRVAGVFAHRHQVD
TDGFLTQVRDPSVRSSELNVLLQALNGRSQVSSQAQLNGLAGAPAREAP"
Query Match 80.8%; Score 20.2; DB 1; Length 346274;
Best Local Similarity 88.0%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CCGCGCGCGGAGAGCCAGCAT 25
Db 57840 CCGCGCGCGCGCGCCAGCAT 57816

RESULT 9
BX640430/c
LOCUS
DEFINITION
Bordetella parapertussis strain 12822, complete genome; Segment
8/14.
ACCESSION
BX640430 BX470249
VERSION
BX640430.1 GI:33566341
KEYWORDS
complete genome.
SOURCE
Bordetella parapertussis
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE
1
Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdeno-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N.,
Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Doggett,J., Feltwell,T., Gobie,A., Hamlin,N., Hauser,H.,
Hollroyd,S., Jagels,K., Leather,S., Moule,S., Norbertczak,H.,
O'Neill,S., Omond,D., Price,C., Rabinowitsch,E., Rutter,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skelton,J., Squares,R., Squares,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrell,D.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
Online Publication
Nat. Genet. DOI:10.1038/NG1227-10.1038/NG1227
2 (bases 1 to 348014)
Sebahia,M.
Direct Submission
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: mes@sanger.ac.uk
FEATURES
source
1. 348014
/organism="Bordetella parapertussis"
/mol_type="genomic DNA"
/strain="12822"
/db_xref="taxon:519"
365..865
/locus_tag="BPP2288"
365..865
/locus_tag="BPP2288"
/notes="ortholog of Bordetella pertussis (BX470248) BP1915"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE37586.1"
/db_xref="GI:33566342"
/translation="MKQHESGIADVVPDGSDFWQVPANGVVRVTLNANTGASV
DFSAGTQVAPGCHVREHFDREVFITETGTALIDGVEHPLAGACPLGKSRK
HSFLNAGPEPLSFFMLPGGLDDFFRIGARQGPAPQFPAPFADRIEANTVF
GWTAKS"
641..864
/locus_tag="BPP2288"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)."
907..1221
/locus_tag="BPP2289"
907..1221
/locus_tag="BPP2289"
/notes="ortholog of Bordetella pertussis (BX470248) BP1916"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE37587.1"
/db_xref="GI:33566343"
/translation="MFIERYRIKPGKAGLVATYEAGLQKLYLGRFLGPFVMSF
IGELNHVVAWYADSLDERDAARSLADPQWQAYLERVTDLIDLQQTRLKPFVSFP
IQ"
1231..2172
/locus_tag="BPP2290"
1231..2172
/locus_tag="BPP2290"
/notes="ortholog of Bordetella pertussis (BX470248) BP1917"
/codon_start=1
/transl_table=11
/product="putative 3-hydroxyacyl-CoA dehydrogenase"
/protein_id="CAE37588.1"
/db_xref="GI:33566344"
/translation="MNHVAVIGGGIIGASVAVFARRGLEVTIVERRDAACLAGLPAEL
AGMERSASLILGAGEPGDVAARICATDAAAVAGRDYQVQVSENLAKRTLFAEL
VLAPAHALLASSTYTGASQFTALAGARCLVAHPMTPHLSPVVEMAASAMTDPQ
VLAGEAFMRISIGQHPVRIKEIPGVNLRLQGLLMMFVRIADDDVISPADALIS
QGLGLRWATLPLGLEVDLNAFGIADYLRQYGHIFNDMAAGQLPAPVDALISALDS
AMRAALPLERLEAKRGMKRAIAGVRVVALHDTYGEKP"
1231..1779
/locus_tag="BPP2290"
/notes="HMMpfam hit to PF02737, 3-hydroxyacyl-CoA
dehydrogenase, NAD binding domain, score 5.3e-28"
```



TITLE Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica  
JOURNAL Online Publication  
REMARK Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227  
REFERENCE 2 (bases 1 to 349672)  
AUTHORS Sebahia,M.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-2003) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: ms5@sanger.ac.uk

FEATURES  
source  
1. 349672  
/organism="Bordetella pertussis Tohama I"  
/mol\_type="genomic DNA"  
/strain="Tohama I"  
/db\_xref="taxon:257313"  
222..833  
/locus\_tag="BP2626"  
222..833  
/locus\_tag="BP2626"  
/notes="Similar to several proteins of undefined function e.g. Pseudomonas aeruginosa hypothetical protein Pa3765 Pa3765 TR:Q9HXN0 (EMBL:AE004795) (194 aa) fasta scores: E(): 8.6e-32, 55.026% id in 189 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAE42902.1"  
/db\_xref="GI:33563677"  
translation="MTTPMPTAFAGHALLAAGFLAEVALAARAADGAGAVFIY  
EATGAIQVDVGTGDAIEIVARLPLPTFPVEPEGSPARPRGRGRLGVVARE  
VLLPRHWLQAGQCGASVALRKLVQARRRHAGADALRRGRDAAYTFMSAAGNLP  
GFEERVALYAGDAERLAAMAGWADVPAYALRLAGRADLEA"  
complement(join(830..2461,2465..2998))  
/locus\_tag="BP2627"  
/pseudo  
complement(join(830..2461,2465..2998))  
/locus\_tag="BP2627"  
/notes="Pseudogene. This CDS contains an in-frame TAG stop codon. The sequence has been checked and believed to be correct. Similar to several including: Agrobacterium tumefaciens Agr\_c142lp agr\_c142l TR:AAK86587 (EMBL:AE008011) (848 aa) fasta scores: E(): 2e-26, 29.915% id in 819 aa and Bordetella pertussis putative autotransporter BapA protein TR:Q9F4B4 (EMBL:AJ277632) (903 aa) fasta scores: E(): 0.059, 23.692% id in 688 aa"  
/pseudo  
/codon\_start=1  
/transl\_table=11  
/product="autotransporter (pseudogene)"  
/db\_xref="PSEUDO:CAE42903.1"  
complement(2462..2464)  
/locus\_tag="BP2627"  
/notes="in frame TAG stop codon in pertussis; CAG in parapertussis and bronchiseptica"  
complement(2897..2998)  
/locus\_tag="BP2627"  
/notes="Signal peptide predicted for BP2628 by SignalP 2.0 HMM (signal peptide probability 0.943) with cleavage site probability 0.901 between residues 34 and 35"  
/pseudo  
complement(2915..2947)  
/locus\_tag="BP2627"  
/notes="PS00013 Prokaryotic membrane lipoprotein lipid attachment site."  
/pseudo  
complement(3248..4346)  
/locus\_tag="BP2629"  
complement(join(3248..3868,3867..4346))  
/locus\_tag="BP2629"  
/notes="Pseudogene. This CDS appears to have a frameshift mutation following codon 187. The frameshift occurs within

a polymeric tract of (GC)3. The sequence has been checked and believed to be correct. Weakly similar to Rhizobium loti hypothetical protein M19176 TR:Q981Z1 (EMBL:AP003015) (386 aa) fasta scores: E(): 4.8e-05, 28% id in 300 aa"  
/pseudo  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein (pseudogene)"  
/db\_xref="PSEUDO:CAE42904.1"  
complement(3868..3873)  
/locus\_tag="BP2629"  
/notes="(GC)3 in pertussis; (GC)4 in parapertussis and bronchiseptica"  
complement(4580..5631)  
/locus\_tag="BP2630"  
complement(4580..5530)  
/locus\_tag="BP2630"  
complement(4580..5530)  
/locus\_tag="BP2630"  
/notes="transposase for IS481 element"  
/codon\_start=1  
/transl\_table=11  
/product="transposase"  
/protein\_id="CAE42905.1"  
/db\_xref="GI:33563678"  
translation="MTHKHARLFLRRLEWQQLIAHQCVCPEARAYGVTAFTVRK  
WLGFLAQGAGLADASSRPTVSPRAIPAKALAIVELRRKRLTOARIAQALGVSAST  
VSVRLARAGSHLADLPAPSPVRYEHQAPGDHLHIDIKLGIQPCRHVTGNERDT  
VEGAWDFVFAIDDHARVAFDTDHPDERPSAVOFKDAVAYVQRLGVTIQRLLTND  
GSAPSRFAALCHLGHKRFTRPYQPTNGKAERFIQSALREWAYHTYQNSQCHRA  
DAMKSWLHYNHWRPHQIGRAVPSISLNDDEVNLLTVHT"  
4580..4611  
/notes="Inverted repeat for IS481 element"  
/rpt\_type="INVERTED"  
complement(4616..5149)  
/locus\_tag="BP2630"  
/notes="HMPFiam hit to PF00665, Integrase core domain"  
complement(5207..5272)  
/locus\_tag="BP2630"  
/notes="Predicted helix-turn-helix motif with score 2406 (+7.38 SD) at aa 87-108, sequence LTQARIAQALGVSASTVSRVLA"  
complement(5601..5631)  
/notes="Inverted repeat for IS481 element"  
/rpt\_type="INVERTED"  
5633..5875  
/locus\_tag="BP2631"  
/pseudo  
5633..5875  
/locus\_tag="BP2631"  
/notes="Pseudogene. This CDS is the C-terminal region of BP1690 (disrupted by the insertion of IS481 element)."  
/pseudo  
/codon\_start=1  
/transl\_table=11  
/product="C-terminal region of a putative exported protein (pseudogene)."  
complement(5946..8582)  
/locus\_tag="BP2632"  
complement(5946..8582)  
/locus\_tag="BP2632"  
/notes="Similar to several Prokaryotic and Eukaryotic proteins of undefined function e.g. Mycobacterium leprae hypothetical protein M1120 TR:Q9C83 (EMBL:AL583920) (873 aa) fasta scores: E(): 7.6e-19, 28.02% id in 910 aa and Leishmania major hypothetical protein L6202.3 TR:Q9NKT9 (EMBL:AC005802) (2354 aa) fasta scores: E(): 1.5e-10, 31.59% id in 652 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative GTP-binding protein"  
/protein\_id="CAE42907.1"  
/db\_xref="GI:33563679"  
translation="MKLRRIALEAFKRPQFVALEDLGGLLIIVGPNAGKSTFVA

variation  
repeat\_unit  
gene  
CDS  
misc\_feature  
misc\_feature  
repeat\_unit  
gene  
CDS  
gene  
CDS  
misc\_feature  
misc\_feature  
gene  
CDS









```

SEQUENCE, 9 unordered pieces.
AC135355
VERSION AC135355.1 GI:23915587
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 218467)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 218467)
McPherson, J.D. and Waterston, R.H.
Direct Submission
JOURNAL Submitted (12-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BB0495F22
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator; 100%
Assembly program: Phrap; version 0.990319
Consensus quality: 215131 bases at least Q40
Consensus quality: 215759 bases at least Q30
Consensus quality: 216077 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1224: contig of 1224 bp in length
1225 1324: gap of unknown length
1325 1329: contig of 1605 bp in length
2930 3029: gap of unknown length
3030 3030: contig of 3485 bp in length
6515 6516: gap of unknown length
6516 19263: contig of 12648 bp in length
19264 19363: gap of unknown length
19364 30033: contig of 10669 bp in length
30033 30133: gap of unknown length
30133 47879: contig of 17747 bp in length
47880 47979: gap of unknown length
47980 86517: contig of 38538 bp in length
86518 135034: contig of 48417 bp in length
135035 135134: gap of unknown length
135135 218467: contig of 83333 bp in length.
FEATURES
source
1..218467
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-495F22"
1..1224
/notes="assembly_name:Contig26"
1325..2929
/notes="assembly_name:Contig27"
3030..6515
misc_feature
misc_feature
misc_feature

```

```

/misc_feature
616..19263
/notes="assembly_name:Contig28"
/misc_feature
19364..30032
/notes="assembly_name:Contig30"
30133..47879
/notes="assembly_name:Contig31"
47980..86517
/notes="assembly_name:Contig32"
86618..135034
/notes="assembly_name:Contig33"
135135..218467
/notes="assembly_name:Contig34"
ORIGIN
Query Match 79.2%; Score 19.8; DB 2; Length 218467;
Best Local Similarity 91.3%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCGCGCGCGCAGCCACACAGG 23
Db 11904 CCGCTGTGCGAGCCACACAGG 11926
RESULT 14
LOCUS AC145568 226332 bp DNA linear HTG 16-OCT-2003
DEFINITION Mus musculus chromosome UNK clone RP24-473H20, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION AC145568
VERSION AC145568.2 GI:37693711
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Wilson, R.K.
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 226332)
Wilson, R.K.
Direct Submission
JOURNAL Submitted (19-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 226332)
Wilson, R.K.
Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
On Oct 16, 2003 this sequence version replaced gi:32996905.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BB0473H20
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator; 100%
Assembly program: Phrap; version 0.990319
Consensus quality: 221614 bases at least Q40
Consensus quality: 221869 bases at least Q30
Consensus quality: 222091 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 1070: contig of 1070 bp in length  
 \* 1071 1170: gap of unknown length  
 \* 1171 4030: contig of 2860 bp in length  
 \* 4031 4130: gap of unknown length  
 \* 4131 7600: contig of 3470 bp in length  
 \* 7601 7700: gap of unknown length  
 \* 7701 19247: contig of 11547 bp in length  
 \* 19248 19347: gap of unknown length  
 \* 19348 31995: contig of 12648 bp in length  
 \* 31996 32095: gap of unknown length  
 \* 32096 43859: contig of 11764 bp in length  
 \* 43860 43959: gap of unknown length  
 \* 43960 58105: contig of 14146 bp in length  
 \* 58106 58205: gap of unknown length  
 \* 58206 73638: contig of 15433 bp in length  
 \* 73639 73738: gap of unknown length  
 \* 73739 108016: contig of 34278 bp in length  
 \* 108017 108117: gap of unknown length  
 \* 108118 138559: contig of 30443 bp in length  
 \* 138560 138659: gap of unknown length  
 \* 138660 178535: contig of 39876 bp in length  
 \* 178536 178635: gap of unknown length  
 \* 178636 226332: contig of 47697 bp in length.

## FEATURES

## source

1. 226332  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP24-473H20"  
 1. 1070  
 /note="assembly\_name:Contig30"  
 1171. 4030  
 /note="assembly\_name:Contig32"  
 4131. 7600  
 /note="assembly\_name:Contig33"  
 7701. 19247  
 /note="assembly\_name:Contig34"  
 19348. 31995  
 /note="assembly\_name:Contig35"  
 32096. 43859  
 /note="assembly\_name:Contig36"  
 43960. 58105  
 /note="assembly\_name:Contig37"  
 58206. 73638  
 /note="assembly\_name:Contig38"  
 73739. 108016  
 /note="assembly\_name:Contig39"  
 108117. 138559  
 /note="assembly\_name:Contig40"  
 138660. 178535  
 /note="assembly\_name:Contig41"  
 178636. 226332  
 /note="assembly\_name:Contig42"

## ORIGIN

Query Match 79.28; Score 19.8; DB 2; Length 226332;  
 Best Local Similarity 91.33; Pred. No. 5.6e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGCGGGGCGAGACCAACAGG 23  
 |||||  
 Db 24636 CCGGTGTGGCAGACCAACAGG 24658

## RESULT 15

AC123190/c

## LOCUS

AC123190 231638 bp DNA linear HTG 08-OCT-2002

## DEFINITION

Rattus norvegicus clone CH230-142A15, \*\*\* SEQUENCING IN PROGRESS

AC123190.3 GI:22857200

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

## AUTHORS

1 (bases 1 to 231638)  
 Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Burch, P., Burrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M.,  
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorensuhewa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwaokemele, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K.,  
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,  
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,  
 Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstock, G. and Gibbs, R. A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
 On Sep 14, 2002 this sequence version replaced gi:21671714.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNKN
Center clone name: CH230-142A15
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 211769 bases at least Q40
Consensus quality: 213755 bases at least Q30
Consensus quality: 215000 bases at least Q20
Estimated insert size: 227243; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 230530: contig of 230530 bp in length
* 230531 230630: gap of unknown length
* 230631 231638: contig of 1008 bp in length.
FEATURES
    source
        1..231638
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-142A15"
            misc_feature
                1..1225
                    /note="wgs_contig"
ORIGIN

```

```

Query Match      79.2%; Score 19.8; DB 2; Length 231638;
Best Local Similarity 91.3%; Pred.No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGCGCGGAGAGCCCAACGAG 23
    |||
Db 70704 CCAGTGGCGGAGAGCCCAACGAG 70682

Search completed: April 29, 2004, 06:02:04
Job time : 383.555 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 1971.27 Seconds

(without alignments)  
378.717 Million cell updates/sec

Title: US-10-624-714-10

Perfect score: 25  
Sequence: 1 ccgcggcgcagagcaaccaggat 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.4	81.6	771	14	CB904869
C 2	20.4	81.6	771	14	CF876097
C 3	19.2	76.8	388	9	AJ559459
C 4	19.2	76.8	541	14	CF361438

5	19.2	76.8	585	28	BH279871
6	19.2	76.8	639	12	BI400600
7	19.2	76.8	956	29	CNS03EXI
8	19.2	76.8	1084	10	BE889974
9	18.8	75.2	259	14	CB927837
10	18.8	75.2	498	14	CD923077
11	18.8	75.2	510	14	CD382617
12	18.8	75.2	560	14	CD382543
13	18.8	75.2	681	14	CF872953
14	18.8	75.2	694	14	CF882304
15	18.8	75.2	705	13	BQ540430
16	18.8	75.2	706	14	CF866131
17	18.8	75.2	739	14	CF865887
18	18.8	75.2	739	14	CF865887
19	18.8	75.2	761	14	CB896092
20	18.8	75.2	761	14	CB902977
21	18.8	75.2	769	14	CF870419
22	18.8	75.2	795	14	CB899733
23	18.8	75.2	797	14	CB904955
24	18.8	75.2	797	14	CF876186
25	18.8	75.2	812	14	CD381788
26	18.8	75.2	823	14	CB900601
27	18.8	75.2	876	14	CD380970
28	18.8	75.2	887	14	CD381549
29	18.8	75.2	893	14	CD381206
30	18.8	75.2	895	14	CD382672
31	18.8	75.2	897	14	CD377270
32	18.8	75.2	970	14	CF866215
33	18.8	75.2	1113	12	EG849338
34	18.8	75.2	1316	29	AG044209
35	18.6	74.4	313	9	AI470955
36	18.6	74.4	364	13	BQ102414
37	18.6	74.4	395	12	BM717846
38	18.6	74.4	416	10	BF514498
39	18.6	74.4	419	13	EX876324
40	18.6	74.4	438	9	AW028722
41	18.6	74.4	444	10	BF115881
42	18.6	74.4	452	9	AI184689
43	18.6	74.4	452	10	BF434279
44	18.6	74.4	452	10	BF061630
45	18.6	74.4	465	9	AL702713

#### ALIGNMENTS

RESULT 1  
CB904869/c  
LOCUS  
DEFINITION  
tric040xd06 T.reesei mycelial culture, Version 3 apr11 Hypocrea  
jecorina cDNA clone tric040xd06, mRNA sequence.  
ACCESSION  
CB904869  
VERSION  
CB904869.1 GI:30119527  
KEYWORDS  
EST.  
SOURCE  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocrea; Hypocrea.  
REFERENCE  
1 (bases 1 to 771)  
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.  
Transcriptional regulation of biomass-degrading enzymes in the  
filamentous fungus Trichoderma reesei  
J. Biol. Chem. 278 (34), 31988-31997 (2003)  
JOURNAL  
MEDLINE  
22803314  
PUBMED  
12788920  
COMMENT  
Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817



```

source
1. 541
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="WARC 3P1G"
/notes="vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

Query Match
Best Local Similarity 87.5%; Pred. No. 7.5e+03; Length 541;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
2 CGCGGGCGGAGAGCCAGCAACAGGAT 25
|||||
241 CGCTGGCGGAGAGCCAGCGGGAT 264

RESULT 5
BH279871
LOCUS
DEFINITION
CH230-178L15, TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-178L15, genomic survey sequence.
ACCESSION
BH279871
VERSION
BH279871.1 GI:17192273
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 585)
Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
JOURNAL
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 178 row: L column: 15
Seq primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers
source
1. 585
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-178L15"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/notes="vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by
Pieter de Jong"

Query Match
Best Local Similarity 87.5%; Pred. No. 7.5e+03; Length 585;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
1. 541
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="WARC 3P1G"
/notes="vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

Query Match
Best Local Similarity 87.5%; Pred. No. 7.5e+03; Length 541;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
2 CGCGGGCGGAGAGCCAGCAACAGGAT 25
|||||
241 CGCTGGCGGAGAGCCAGCGGGAT 264

RESULT 6
BH400600/c
LOCUS
DEFINITION
MI-P-AV1-ngf-c-11-0-UI-s1 MI-P-AV1 Sus scrofa cDNA clone
MI-P-AV1-ngf-c-11-0-UI 3', mRNA sequence.
ACCESSION
BH400600
VERSION
BH400600.1 GI:15179661
KEYWORDS
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 639)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
97044477
MEDLINE
889548
PUBMED
Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized placenta library cDNA Library Preparation: M.B. Soares
Lab, University of Iowa EST sequencing: M.B. Soares Lab, University
of Iowa Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Location/Qualifiers
source
1. 639
/organism="Sus scrofa"
/mol_type="mRNA"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-AV1-ngf-c-11-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="MI-P-AV1"
/notes="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; The MI-P-AV1
library is normalized library derived from the MI-P-AV0
library, ultimately derived from placenta tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
http://pigest.genome.iastate.edu/. The procedure used to
create this library has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG TISSUE=placenta
TAG LIB=MI-P-AV1
TAG_SEQ=ATTGG"

ORIGIN
Query Match
Best Local Similarity 87.5%; Pred. No. 7.8e+03; Length 639;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
1 CGCGGGCGGAGAGCCAGCAACAGGAT 24
|||||
527 CGCGGGCGGAGAGCCAGCAACAGGAT 550

BEST LOCAL SIMILARITY 87.5%; Pred. No. 7.6e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN
1 CGCGGGCGGAGAGCCAGCAACAGGAT 24
|||||
527 CGCGGGCGGAGAGCCAGCAACAGGAT 550

```

Qy 2 CGCGGGCGGACAGCCACACAGGAT 25  
 Db 626 CGTGGCGGACAGCCAGCGGAT 603

RESULT 7  
 CNS03EXI 956 bp DNA linear GSS 01-SEP-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
 021A19 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.

ACCESSION AL240975  
 VERSION AL240975.1 GI:7961744  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE 1  
 AUTHORS Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,  
 Bernot A., Fizames C., Wincker P., Brottier P., Quetier P.,  
 Saurin W. and Weissenbach J.  
 TITLE Estimate of human gene number provided by genome-wide analysis  
 using Tetraodon nigroviridis DNA sequence  
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)  
 MEDLINE 20296633  
 PUBMED 10835645

REFERENCE 2  
 AUTHORS Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C.,  
 Fizames C., Fischer C., Bouneau L., Billault A., Quetier P.,  
 Saurin W., Bernot A. and Weissenbach J.  
 TITLE Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Genome Res. 10 (7), 939-949 (2000)  
 MEDLINE 20359837  
 PUBMED 10899143

REFERENCE 3 (bases 1 to 956)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.  
 FEATURES  
 source  
 Location/Qualifiers  
 1. 956  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="021A19"  
 /clone\_lib="G"  
 /notes="Genoscope sequence ID : C0BG021AA10SP1-end :  
 PUC-Ori"

ORIGIN  
 Query Match 76.8%; Score 19.2; DB 29; Length 956;  
 Best Local Similarity 87.5%; Pred. No. 8.6e+03;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGCGGGCGGACAGCCACACAGGAT 25  
 Db 321 CGCGGGCGGACAGCCACCTGGT 344

RESULT 8  
 BE889974 1084 bp mRNA linear EST 20-OCT-2000  
 LOCUS 601512234F1 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:3913664 5',  
 DEFINITION mRNA sequence.

ACCESSION BE889974  
 VERSION BE889974.1 GI:10347833  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1084)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLA9734 row: c column: 09  
 High quality sequence stop: 436.  
 Location/Qualifiers  
 1. 1084  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3913664"  
 /tissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_71"  
 /notes="Torsan; uterus; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.1 Kb."

ORIGIN  
 Query Match 76.8%; Score 19.2; DB 10; Length 1084;  
 Best Local Similarity 87.5%; Pred. No. 8.9e+03;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGCGGGCGGACAGCCACACAGGA 24  
 Db 1022 CGCGGGCGGACAGCCACACGCA 1045

RESULT 9  
 CB927837 259 bp mRNA linear EST 28-APR-2003  
 LOCUS ABAL\_34\_H08\_gl\_A012 Abscisic acid-treated seedlings Sorghum bicolor  
 DEFINITION cDNA clone ABAL\_34\_H08\_A012 5', mRNA sequence.

ACCESSION CB927837  
 VERSION CB927837.1 GI:30164108  
 KEYWORDS EST.  
 SOURCE Sorghum bicolor (sorghum)  
 ORGANISM Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 259)  
 AUTHORS Cordonnier-Pratt M.-M., Wentzel V., Suzuki Y., Sugano S.,  
 Klein R.R., Liang C., Sun F., Sullivan R., Shah M., Buchanan C.D.,  
 Eastman A. and Pratt L.H.  
 TITLE An EST database from Sorghum: ABAL-treated seedlings  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmpratt@uga.edu  
 Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
 the Human Genome Center, University of Tokyo Institute of Medical



Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTGTGCTCTAAAGCTGGC).

#### FEATURES

Location/Qualifiers  
1..259  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cultiVar="IS1620C"  
/db\_xref="taxon:4558"  
/clone="ABAL\_34\_H09\_A012"  
/lab\_host="DHI08-T1 phage-resistant E. coli"  
/clone\_lib="Abcisic acid-treated seedlings"  
/notes="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. After 12 days, medium was supplemented with 1 mM abscisic acid (ABA), while leaves were misted with a solution of 1 mM ABA. Roots and leaves were harvested after 3, 6, 12, and 24 hr and material from all time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCACTGTG). XhoI excises the cDNA insert."

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 259;  
Best Local Similarity 90.9%; Pred. No. 8.6e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGCGCGGAGAGCCGACAGG 22  
DB 55 CCGCGCGCGGAGAGCCGACAGG 76

#### RESULT 10

CD923077 498 bp mRNA linear EST 15-JUL-2003  
LOCUS G750.106G21P010531 G750 Triticum aestivum cDNA clone G750106G21, mRNA sequence.  
DEFINITION CD923077

ACCESSION CD923077.1 GI:32770841

VERSION EST.

KEYWORDS Triticum aestivum (bread wheat)

SOURCE Triticum aestivum

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

1 (bases 1 to 498)

#### REFERENCE

Genoplante, a major partnership french program in plant genomics  
Unpublished (2003)

Contact: Genoplante

Genoplante

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french plant genomics programme 'Genoplante' (<http://www.genoplante.com>) and <http://genoplante-info.infobiogen.fr>.

#### FEATURES

Location/Qualifiers  
1..498  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultiVar="recital"  
/db\_xref="taxon:4565"  
/clone="G750106G21"  
/tissue\_type="grain (750 degrees per day after pollination)"  
/clone\_lib="G750"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 498;  
Best Local Similarity 90.9%; Pred. No. 1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCGCGCGGAGAGCCGACAGG 22  
DB 412 CCGCGCGCGGAGAGCCGACAGG 433

#### RESULT 11

CD382617 510 bp mRNA linear EST 31-MAY-2003  
LOCUS PTMM07787 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum tricornutum cDNA 5', mRNA sequence.  
DEFINITION CD382617

ACCESSION CD382617.1 GI:31258231

VERSION EST.

KEYWORDS Phaeodactylum tricornutum

SOURCE Phaeodactylum tricornutum

ORGANISM

Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae; Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.

1 (bases 1 to 510)

AUTHORS Scala, S., Carels, N., Falciatore, A., Chiusano, M.L. and Bowler, C.

TITLE Genome properties of the diatom Phaeodactylum tricornutum

JOURNAL Plant Physiol.

MEDLINE 22111123

PUBMED 12114555

COMMENT Contact: Bowler C

Laboratory of Molecular Plant Biology

Stazione Zoologica 'Anton Dohrn', Italy

Villa Comunale, I-80121, Napoli, Italy

Tel: 39 081 583 3268/3211

Fax: 39 081 764 1355

Email: [chris@alpha.szn.it](mailto:chris@alpha.szn.it)

Diatom EST Database (<http://avesthagen.sznbowler.com>)

Seq primer: T3 backward

POLYA=Yes.

Location/Qualifiers

1..510

/organism="Phaeodactylum tricornutum"

/mol\_type="mRNA"

/db\_xref="taxon:2850"

/cell\_line="CCMP632"

/clone\_lib="Phaeodactylum tricornutum Uni-Zap XR"

/note="Vector: Uni-Zap XR vector; Site\_1: Eco RI; Site\_2: Xho I"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 510;  
Best Local Similarity 90.9%; Pred. No. 1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCGCGCGGAGAGCCGACAGG 24  
DB 124 GCGCGCGGAGAGCCGACAGG 103

#### RESULT 12

CD382543 560 bp mRNA linear EST 31-MAY-2003  
LOCUS PTMM07713 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum tricornutum cDNA 5', mRNA sequence.  
DEFINITION CD382543

ACCESSION CD382543.1 GI:31258157

VERSION EST.

KEYWORDS Phaeodactylum tricornutum

SOURCE Phaeodactylum tricornutum

ORGANISM

Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae; Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.

1 (bases 1 to 560)

AUTHORS Scala, S., Carels, N., Falciatore, A., Chiusano, M.L. and Bowler, C.

TITLE Genome properties of the diatom *Phaeodactylum tricornutum*  
JOURNAL Plant Physiol. 129 (3), 993-1002 (2002)  
MEDLINE 22111123  
PUBMED 12114555

COMMENT Contact: Bowler C  
Laboratory of Molecular Plant Biology  
Stazione Zoologica 'Anton Dohrn'  
Villa Comunale, I-80121, Napoli, Italy  
Tel: 39 081 583 3268/3211  
Fax: 39 081 764 1355

Email: chris@alpha.szn.it  
Diatom EST Database (<http://avesthagen.sznbowler.com>)  
Seq primer: T3 backward  
POLYA=yes.

FEATURES  
source Location/Qualifiers  
1. .560  
/organism="Phaeodactylum tricornutum"  
/mol\_type="mRNA"  
/db\_xref="taxon:2850"  
/cell\_line="COP632"  
/clone\_lib="Phaeodactylum tricornutum Uni-Zap XR"  
/notes="vector: Uni-Zap XR vector; Site\_1: Eco RI; Site\_2: Xho I"

ORIGIN  
Query Match 75.2%; Score 18.8; DB 14; Length 560;  
Best Local Similarity 90.9%; Pred. No. 1.1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CGCGCGCGCAGACCAACGAG 24  
|||||  
Db 112 CGCGCGCGCATCCACGAG 91

RESULT 13  
CF872953/c  
LOCUS  
DEFINITION trico32xh04.b1 T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone trico32xh04, mRNA sequence.

ACCESSION CF872953  
VERSION CF872953.1 GI:38127635  
KEYWORDS EST.

SOURCE  
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 681)  
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., and Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

TITLE Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset

JOURNAL  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: IT-F1 primer.

FEATURES  
source Location/Qualifiers  
1. .681  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico32xh04"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/notes="vector: PREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN  
Query Match 75.2%; Score 18.8; DB 14; Length 681;  
Best Local Similarity 90.9%; Pred. No. 1.1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGCGCGCGCAGACCAACGAG 23  
|||||  
Db 192 CGCGCGCGCATGCCACGAG 171

RESULT 14  
CF882304/c  
LOCUS  
DEFINITION trico086xg05.b1 T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone trico086xg05, mRNA sequence.

ACCESSION CF882304  
VERSION CF882304.1 GI:38136986  
KEYWORDS EST.

SOURCE  
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 694)  
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

TITLE Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset

JOURNAL  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: IT-F1 primer.

FEATURES  
source Location/Qualifiers  
1. .694  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico086xg05"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"

/note="vector: PREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 694;  
Best Local Similarity 90.9%; Pred. No. 1.1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGCGCGCGCAGACCAACGAG 23  
|||||  
Db 156 CGCGCGCGCATGCCACGAG 135

RESULT 15  
BQ540430/c  
LOCUS  
DEFINITION PTAM0748 Phaeodactylum tricornutum Uni-Zap XR Phaeodactylum tricornutum cDNA 5', mRNA sequence.

ACCESSION BQ540430  
VERSION BQ540430.1 GI:21396000  
KEYWORDS EST.

SOURCE  
ORGANISM Phaeodactylum tricornutum  
Phaeodactylum tricornutum

Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;  
Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.

REFERENCE

1 (bases 1 to 705)  
Scala, S., Carels, N., Falcione, A., Chiusano, M.L. and Bowler, C.  
Genome properties of the diatom *Phaeodactylum tricornutum*  
Plant Physiol. 129 (3), 993-1002 (2002)  
22111123  
PUBMED  
12114555  
Contact: Bowler C  
Laboratory of Molecular Plant Biology  
Stazione Zoologica 'Anton Dohrn'  
Villa Comunale, I-80121, Napoli, Italy  
Tel: 39 081 583 3268/3211  
Fax: 39 081 764 1355  
Email: chris@alpha.szn.it  
Seq primer: T3 backward.

FEATURES

source  
1..705  
/organism="Phaeodactylum tricornutum"  
/mol\_type="mRNA"  
/db\_xref="taxon:2850"  
/cell\_line="CCMP632"  
/clone\_lib="Phaeodactylum tricornutum Uni-Zap XR"  
/notes="Vector: Uni-Zap XR vector; Site\_1: Eco RI; Site\_2:  
Xho I"

ORIGIN

Query Match 75.2%; Score 18.8; DB 13; Length 705;  
Best Local Similarity 90.9%; Pred. No. 1.1e+04;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GCGGCGGCGAGCCACCAGGA 24  
|||||  
Db 139 GCGGCGGCGGATCCACCAGGA 118

Search completed: April 29, 2004, 11:37:16  
Job time : 1974.27 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 44.1176 Seconds  
(without alignments)  
314.472 Million cell updates/sec

Title: US-10-624-714-10  
Perfect score: 25  
Sequence: 1 CCGCGGCGGCGAGCCACCAGGAT 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCRUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	25	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	18.2	72.8	534	4 US-09-252-991A-5352	Sequence 5352, Ap
C 4	18.2	72.8	1554	4 US-09-252-991A-5191	Sequence 5191, Ap
C 5	18.2	72.8	1815	4 US-09-252-991A-5254	Sequence 5254, Ap
C 6	17.6	70.4	651	4 US-09-252-991A-4491	Sequence 4491, Ap
C 7	17.6	70.4	1194	4 US-09-489-039A-5294	Sequence 5294, Ap
C 8	17.6	70.4	1473	4 US-09-252-991A-4196	Sequence 4196, Ap
C 9	17.6	70.4	1778	4 US-08-934-386-4	Sequence 4, Appli
C 10	17.6	70.4	3105	4 US-09-252-991A-4398	Sequence 4398, Ap
C 11	17.6	70.4	11748	1 US-08-611-107-30	Sequence 30, Appl
C 12	17.2	68.8	906	4 US-09-252-991A-5994	Sequence 5994, Ap
C 13	17.2	68.8	1119	4 US-09-252-991A-6391	Sequence 6391, Ap
C 14	17.2	68.8	1224	4 US-09-252-991A-4775	Sequence 4775, Ap
C 15	17.2	68.8	1230	4 US-09-252-991A-4733	Sequence 4733, Ap
C 16	17.2	68.8	1230	4 US-09-252-991A-4767	Sequence 4767, Ap
C 17	17.2	68.8	1356	4 US-09-252-991A-4740	Sequence 4740, Ap
C 18	17.2	68.8	1443	4 US-09-252-991A-6308	Sequence 6308, Ap
C 19	17.2	68.8	2218	2 US-08-985-090-4	Sequence 4, Appli
C 20	17.2	68.8	2218	3 US-09-165-543-31	Sequence 31, Appl
C 21	17.2	68.8	3244	3 US-09-165-543-4	Sequence 4, Appli
C 22	17	68.0	206	4 US-09-833-381-1124	Sequence 1124, Ap
C 23	17	68.0	558	4 US-09-669-751-154	Sequence 154, App
C 24	17	68.0	909	4 US-09-252-991A-1620	Sequence 1620, Ap
C 25	17	68.0	1062	4 US-09-489-039A-919	Sequence 919, App
C 26	17	68.0	1068	4 US-09-252-991A-1546	Sequence 1546, Ap
C 27	17	68.0	1309	3 US-08-362-495-4	Sequence 4, Appli

C 28	17	68.0	1309	4 US-09-408-508-4	Sequence 4, Appli
C 29	17	68.0	1383	4 US-09-489-039A-3348	Sequence 3348, Ap
C 30	17	68.0	1437	4 US-09-489-039A-731	Sequence 731, App
C 31	17	68.0	2016	4 US-09-252-991A-1449	Sequence 1449, Ap
C 32	17	68.0	2022	4 US-09-252-991A-1573	Sequence 1573, Ap
C 33	17	68.0	4325	2 US-08-888-497-21	Sequence 21, Appl
C 34	17	68.0	4325	4 US-09-362-230-21	Sequence 21, Appl
C 35	17	68.0	4325	5 PCT-US94-07926-21	Sequence 21, Appl
C 36	17	68.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 37	17	68.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 38	16.6	66.4	47	4 US-09-422-978-3425	Sequence 3425, Ap
C 39	16.6	66.4	242	2 US-08-673-190A-7	Sequence 7, Appli
C 40	16.6	66.4	278	4 US-09-313-294A-2824	Sequence 2824, Ap
C 41	16.6	66.4	638	4 US-09-449-218D-11	Sequence 11, Appl
C 42	16.6	66.4	638	4 US-09-668-529A-11	Sequence 11, Appl
C 43	16.6	66.4	638	4 US-09-668-037A-11	Sequence 11, Appl
C 44	16.6	66.4	642	4 US-09-449-218D-45	Sequence 45, Appl
C 45	16.6	66.4	642	4 US-09-668-529A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 100.0%; Score 25; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGCGGCGGCGAGCCACCAGGAT 25  
Db 1113158 CCGCGGCGGCGAGCCACCAGGAT 1113134

RESULT 2  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 25; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGACCAACAGGAT 25
Db 1113135 CCGCGCGCGCAGACCAACAGGAT 1113111

RESULT 3
US-09-252-991A-5352/c
; Sequence 5352, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5352
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5352

Query Match      72.8%; Score 18.2; DB 4; Length 534;
Best Local Similarity 87.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGCGCGCGCAGACCAACAGGAT 24
Db 193 CCGCGCGCGCAGACCAACAGGAT 171

RESULT 4
US-09-252-991A-5191
; Sequence 5191, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5191
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5191

Query Match      72.8%; Score 18.2; DB 4; Length 1554;
Best Local Similarity 87.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 25; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGACCAACAGGAT 25
Db 1113135 CCGCGCGCGCAGACCAACAGGAT 1113111

RESULT 3
US-09-252-991A-5352/c
; Sequence 5352, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5352
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5352

Query Match      72.8%; Score 18.2; DB 4; Length 534;
Best Local Similarity 87.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGCGCGCGCAGACCAACAGGAT 24
Db 193 CCGCGCGCGCAGACCAACAGGAT 171

RESULT 4
US-09-252-991A-5191
; Sequence 5191, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5191
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5191

Query Match      72.8%; Score 18.2; DB 4; Length 1554;
Best Local Similarity 87.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 25; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGACCAACAGGAT 25
Db 1113135 CCGCGCGCGCAGACCAACAGGAT 1113111

RESULT 3
US-09-252-991A-5352/c
; Sequence 5352, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5352
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5352

Query Match      72.8%; Score 18.2; DB 4; Length 1815;
Best Local Similarity 87.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGCGCGCGCAGACCAACAGGAT 24
Db 1172 CCGCGCGCGCAGACCAACAGGAT 1194

RESULT 6
US-09-252-991A-4491/c
; Sequence 4491, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4491
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4491

Query Match      70.4%; Score 17.6; DB 4; Length 651;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGCGCGCGCAGACCAACAGGAT 24
Db 526 CCGCGCGCGCAGACCAACAGGAT 503

RESULT 7
US-09-489-039A-5294/c
; Sequence 5294, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
```

```
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5294
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5294

Query Match 70.4%; Score 17.6; DB 4; Length 1194;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCGCGCGGCGAGCCCAACGAGAT 25
Db 801 CCGCGCGGCGAGGCCCATCAGGAT 778

RESULT 8
US-09-252-991A-4196
; Sequence 4196, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4196
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4196

Query Match 70.4%; Score 17.6; DB 4; Length 1473;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGCGCGGCGAGCCCAACGAGA 24
Db 1319 CCGCGCGGCGAGGCCCATCAGCA 1342

RESULT 9
US-08-934-386-4/c
; Sequence 4, Application US/08934386
; Patent No. 6306636
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: Methods for Detecting Nucleic Acid
; TITLE OF INVENTION: Segments Encoding Acetyl-CoA Carboxylase
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 433
; CITY: Houston
; STATE: Texas
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,386
; FILING DATE: 19-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARSB:521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1778 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-934-386-4

Query Match 70.4%; Score 17.6; DB 4; Length 1778;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGCGCGGCGAGCCCAACGAGA 24
Db 558 CCGCGCGGCGAGGCCCATCAGCA 535

RESULT 10
US-09-252-991A-4398/c
; Sequence 4398, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4398
; LENGTH: 3105
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4398

Query Match 70.4%; Score 17.6; DB 4; Length 3105;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGCGCGGCGAGCCCAACGAGA 24
Db 773 CCGCGCGGCGAGGCCCATCAGCA 750

RESULT 11
US-08-611-107-30
; Sequence 30, Application US/08611107
; Patent No. 5801233
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE AND USES
; TITLE OF INVENTION: THEREFOR
```





Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 4733  
LENGTH: 1230  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4733

Query Match 88.8%; Score 17.2; DB 4; Length 1230;  
Best Local Similarity 86.4%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 CGGCGGCGAGCAACCCAGGAT 25  
Db 882 CGGCGGCGAGCAACCCAGGAT 903

Search completed: April 29, 2004, 11:47:01  
Job time : 59.1176 secs

This Page Blank (uspto)